



US011710309B2

(12) **United States Patent**  
**Shotton et al.**

(10) **Patent No.:** **US 11,710,309 B2**  
(45) **Date of Patent:** **Jul. 25, 2023**

(54) **CAMERA/OBJECT POSE FROM PREDICTED COORDINATES**

(58) **Field of Classification Search**  
None  
See application file for complete search history.

(71) Applicant: **Microsoft Technology Licensing, LLC**,  
Redmond, WA (US)

(56) **References Cited**

(72) Inventors: **Jamie Daniel Joseph Shotton**,  
Cambridge (GB); **Benjamin Michael Glocker**,  
Cambridge (GB); **Christopher Zach**, Cambridge (GB);  
**Shahram Izadi**, Cambridge (GB); **Antonio Criminisi**,  
Cambridge (GB); **Andrew William Fitzgibbon**,  
Cambridge (GB)

U.S. PATENT DOCUMENTS

4,215,364 A 7/1980 Takanashi et al.  
4,627,620 A 12/1986 Yang  
(Continued)

FOREIGN PATENT DOCUMENTS

CN 1523352 A 8/2004  
CN 1734499 A 2/2006  
(Continued)

(73) Assignee: **Microsoft Technology Licensing, LLC**,  
Redmond, WA (US)

OTHER PUBLICATIONS

(\*) Notice: Subject to any disclaimer, the term of this  
patent is extended or adjusted under 35  
U.S.C. 154(b) by 0 days.

Lepetit, Vincent, Pascal Lager, and Pascal Fua. "Randomized trees  
for real-time keypoint recognition." 2005 IEEE Computer Society  
Conference on Computer Vision and Pattern Recognition (CVPR'05).  
vol. 2. IEEE, 2005. (Year: 2005).\*

(21) Appl. No.: **15/895,990**

(Continued)

(22) Filed: **Feb. 13, 2018**

*Primary Examiner* — Sean M Conner

(65) **Prior Publication Data**

US 2018/0285697 A1 Oct. 4, 2018

(74) *Attorney, Agent, or Firm* — Barta, Jones & Foley,  
PLLC

**Related U.S. Application Data**

(63) Continuation of application No. 13/774,145, filed on  
Feb. 22, 2013, now Pat. No. 9,940,553.

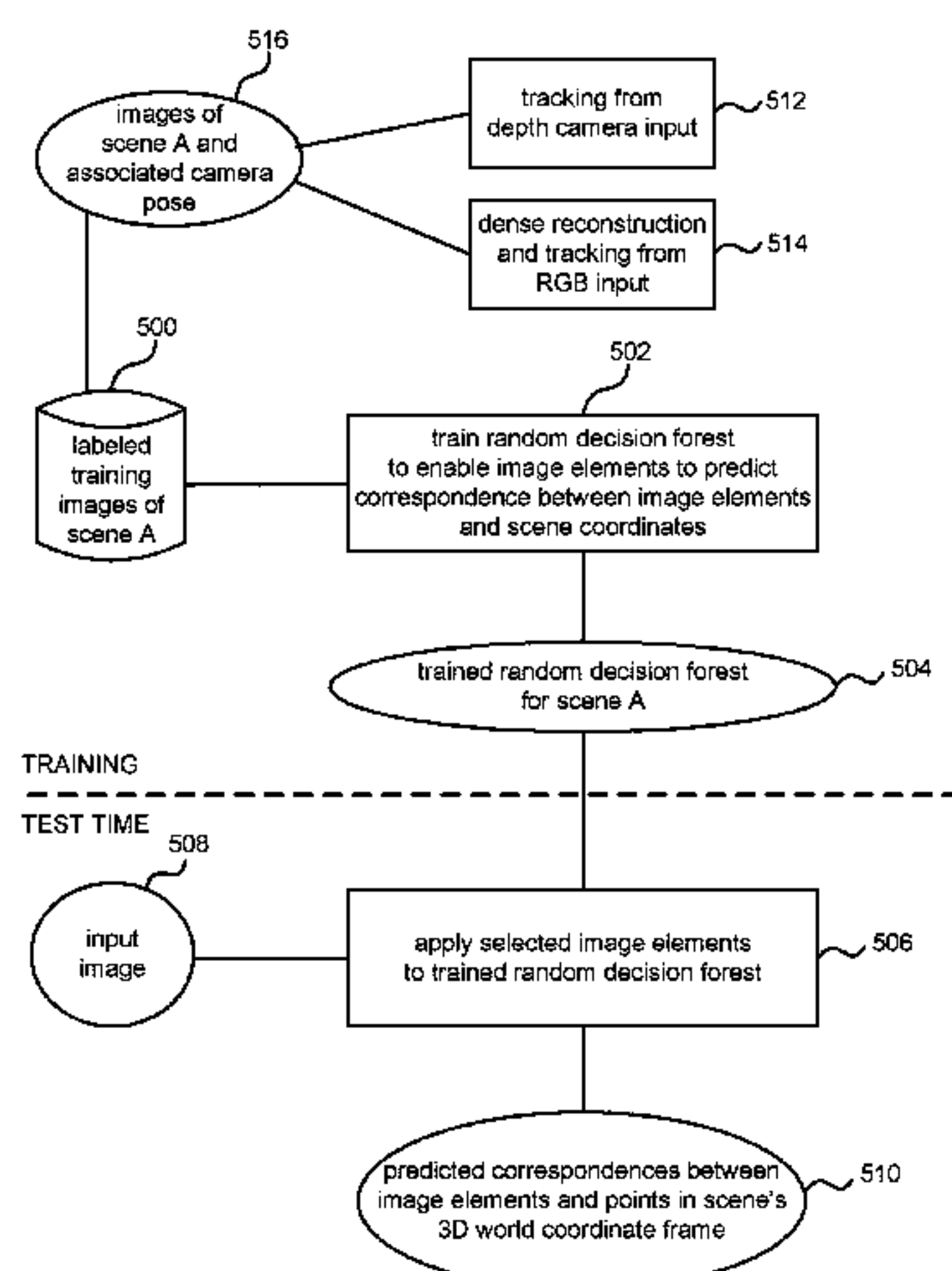
(51) **Int. Cl.**  
**G06V 10/762** (2022.01)  
**G06V 20/20** (2022.01)  
(Continued)

(57) **ABSTRACT**

Camera or object pose calculation is described, for example,  
to relocalize a mobile camera (such as on a smart phone) in  
a known environment or to compute the pose of an object  
moving relative to a fixed camera. The pose information is  
useful for robotics, augmented reality, navigation and other  
applications. In various embodiments where camera pose is  
calculated, a trained machine learning system associates  
image elements from an image of a scene, with points in the  
scene's 3D world coordinate frame. In examples where the  
camera is fixed and the pose of an object is to be calculated,  
the trained machine learning system associates image ele-  
ments from an image of the object with points in an object  
coordinate frame. In examples, the image elements may be

(52) **U.S. Cl.**  
CPC ..... **G06V 20/20** (2022.01); **G06F 18/214**  
(2023.01); **G06F 18/231** (2023.01);  
(Continued)

(Continued)



noisy and incomplete and a pose inference engine calculates an accurate estimate of the pose.

**20 Claims, 10 Drawing Sheets**

- (51) **Int. Cl.**  
*G06N 3/08* (2023.01)  
*G06F 18/214* (2023.01)  
*G06F 18/231* (2023.01)  
*G06F 18/243* (2023.01)  
*G06V 10/764* (2022.01)  
*G06V 10/774* (2022.01)
- (52) **U.S. Cl.**  
 CPC ..... *G06F 18/24323* (2023.01); *G06V 10/764* (2022.01); *G06V 10/7625* (2022.01); *G06V 10/774* (2022.01)

(56) **References Cited**

U.S. PATENT DOCUMENTS

4,630,910 A	12/1986	Ross et al.	5,745,126 A	4/1998	Jain et al.
4,645,458 A	2/1987	Williams	5,801,717 A	9/1998	Engstrom et al.
4,695,953 A	9/1987	Blair et al.	5,852,672 A	12/1998	Lu
4,702,475 A	10/1987	Elstein et al.	5,875,108 A	2/1999	Hoffberg et al.
4,711,543 A	12/1987	Blair et al.	5,877,803 A	3/1999	Wee et al.
4,751,642 A	6/1988	Silva et al.	5,913,727 A	6/1999	Ahdoot
4,796,997 A	1/1989	Svetkoff et al.	5,926,568 A	7/1999	Chaney et al.
4,809,065 A	2/1989	Harris et al.	5,930,379 A	7/1999	Rehg et al.
4,817,950 A	4/1989	Goo	5,930,392 A	7/1999	Ho
4,843,568 A	6/1989	Krueger et al.	5,933,125 A	8/1999	Fernie et al.
4,893,183 A	1/1990	Nayar	5,980,256 A	11/1999	Carmein
4,901,362 A	2/1990	Terzian	5,989,157 A	11/1999	Walton
4,925,189 A	5/1990	Braeunig	5,995,649 A	11/1999	Marugame
5,101,444 A	3/1992	Wilson et al.	6,005,548 A	12/1999	Latypov et al.
5,148,154 A	9/1992	MacKay et al.	6,009,210 A	12/1999	Kang
5,184,295 A	2/1993	Mann	6,009,359 A	12/1999	El-Hakim et al.
5,229,754 A	7/1993	Aoki et al.	6,054,991 A	4/2000	Crane et al.
5,229,756 A	7/1993	Kosugi et al.	6,057,909 A	5/2000	Yahav et al.
5,239,463 A	8/1993	Blair et al.	6,058,205 A	5/2000	Bahl et al.
5,239,464 A	8/1993	Blair et al.	6,066,075 A	5/2000	Poulton
5,288,078 A	2/1994	Capper et al.	6,072,494 A	6/2000	Nguyen
5,295,491 A	3/1994	Gevens	6,073,489 A	6/2000	French et al.
5,320,538 A	6/1994	Baum	6,077,201 A	6/2000	Cheng
5,347,306 A	9/1994	Nitta	6,084,979 A	7/2000	Kanade et al.
5,385,519 A	1/1995	Hsu et al.	6,098,458 A	8/2000	French et al.
5,405,152 A	4/1995	Katanics et al.	6,100,517 A	8/2000	Yahav et al.
5,417,210 A	5/1995	Funda et al.	6,100,896 A	8/2000	Strohecker et al.
5,423,554 A	6/1995	Davis	6,101,289 A	8/2000	Kellner
5,454,043 A	9/1995	Freeman	6,128,003 A	10/2000	Smith et al.
5,469,740 A	11/1995	French et al.	6,130,677 A	10/2000	Kunz
5,495,576 A	2/1996	Ritchey	6,141,463 A	10/2000	Covell et al.
5,516,105 A	5/1996	Eisenbrey et al.	6,147,678 A	11/2000	Kumar et al.
5,524,637 A	6/1996	Erickson	6,147,764 A	11/2000	Handa
5,534,917 A	7/1996	MacDougall	6,152,856 A	11/2000	Studor et al.
5,563,988 A	10/1996	Maes et al.	6,159,100 A	12/2000	Smith
5,577,981 A	11/1996	Jarvik	6,173,066 B1	1/2001	Peurach et al.
5,580,249 A	12/1996	Jacobsen et al.	6,181,343 B1	1/2001	Lyons
5,594,469 A	1/1997	Freeman et al.	6,188,777 B1	2/2001	Darrell et al.
5,597,309 A	1/1997	Riess	6,215,890 B1	4/2001	Matsuo et al.
5,616,078 A	4/1997	Oh	6,215,898 B1	4/2001	Woodfill et al.
5,617,312 A	4/1997	Iura et al.	6,226,396 B1	5/2001	Marugame
5,638,300 A	6/1997	Johnson	6,229,913 B1	5/2001	Nayar et al.
5,641,288 A	6/1997	Zaenglein, Jr.	6,229,918 B1	5/2001	Toyama
5,644,689 A	7/1997	Ban et al.	6,256,033 B1	7/2001	Nguyen
5,682,196 A	10/1997	Freeman	6,256,400 B1	7/2001	Takata et al.
5,682,229 A	10/1997	Wangler	6,283,860 B1	9/2001	Lyons et al.
5,690,582 A	11/1997	Ulrich et al.	6,289,112 B1	9/2001	Jain et al.
5,696,591 A	12/1997	Bilhorn et al.	6,299,308 B1	10/2001	Voronka et al.
5,699,444 A	12/1997	Palm	6,308,565 B1	10/2001	French et al.
5,703,367 A	12/1997	Hashimoto et al.	6,316,934 B1	11/2001	Amorai-Moriya et al.
5,704,837 A	1/1998	Iwasaki et al.	6,363,160 B1	3/2002	Bradski et al.
5,715,834 A	2/1998	Bergamasco et al.	6,384,819 B1	5/2002	Hunter
			6,405,151 B1	6/2002	Fujii et al.
			6,411,744 B1	6/2002	Edwards
			6,430,997 B1	8/2002	French et al.
			6,476,834 B1	11/2002	Doval et al.
			6,496,598 B1	12/2002	Harman
			6,498,628 B2	12/2002	Iwamura
			6,502,515 B2	1/2003	Burckhardt et al.
			6,503,195 B1	1/2003	Keller et al.
			6,504,569 B1	1/2003	Jasinski et al.
			6,512,838 B1	1/2003	Rafii et al.
			6,539,931 B2	4/2003	Trajkovic et al.
			6,570,555 B1	5/2003	Prevost et al.
			6,590,669 B1	7/2003	Wagner
			6,603,880 B2	8/2003	Sakamoto
			6,633,294 B1	10/2003	Rosenthal et al.
			6,640,202 B1	10/2003	Dietz et al.
			6,661,918 B1	12/2003	Gordon et al.
			6,671,049 B1 *	12/2003	Silver ..... G03F 9/7076 356/399
			6,674,877 B1	1/2004	Jojic et al.
			6,681,031 B2	1/2004	Cohen et al.
			6,714,665 B1	3/2004	Hanna et al.
			6,731,799 B1	5/2004	Sun et al.
			6,738,066 B1	5/2004	Nguyen
			6,741,756 B1	5/2004	Toyama et al.
			6,765,726 B2	7/2004	French et al.
			6,771,277 B2	8/2004	Ohba



(56)

## References Cited

## U.S. PATENT DOCUMENTS

6,781,618 B2	8/2004	Beardsley	7,607,509 B2	10/2009	Schmiz et al.
6,788,809 B1	9/2004	Grzeszczuk et al.	7,620,202 B2	11/2009	Fujimura et al.
6,801,637 B2	10/2004	Voronka et al.	7,627,447 B2	12/2009	Marsh et al.
6,823,204 B2	11/2004	Grass et al.	7,648,460 B2	1/2010	Simopoulos et al.
6,850,635 B2	2/2005	Gerard et al.	7,668,340 B2	2/2010	Cohen et al.
6,873,723 B1	3/2005	Aucsmith et al.	7,680,298 B2	3/2010	Roberts et al.
6,876,496 B2	4/2005	French et al.	7,683,954 B2	3/2010	Ichikawa et al.
6,879,709 B2	4/2005	Tian et al.	7,684,592 B2	3/2010	Paul et al.
6,937,742 B2	8/2005	Roberts et al.	7,697,748 B2	4/2010	Dimsdale et al.
6,940,538 B2	9/2005	Rafey et al.	7,701,439 B2	4/2010	Hillis et al.
6,950,534 B2	9/2005	Cohen et al.	7,702,130 B2	4/2010	Im et al.
6,963,338 B1	11/2005	Bachelder et al.	7,704,135 B2	4/2010	Harrison, Jr.
7,003,134 B1	2/2006	Covell et al.	7,707,162 B2	4/2010	Naphade et al.
7,006,236 B2	2/2006	Tomasi et al.	7,710,391 B2	5/2010	Bell et al.
7,034,818 B2	4/2006	Perry et al.	7,728,839 B2	6/2010	Yang et al.
7,035,431 B2	4/2006	Blake et al.	7,729,530 B2	6/2010	Antonov et al.
7,036,094 B1	4/2006	Cohen et al.	7,746,345 B2	6/2010	Hunter
7,038,855 B2	5/2006	French et al.	7,760,182 B2	7/2010	Ahmad et al.
7,039,239 B2	5/2006	Loui et al.	7,809,167 B2	10/2010	Bell
7,039,676 B1	5/2006	Day et al.	7,834,846 B1	11/2010	Bell
7,042,440 B2	5/2006	Pryor et al.	7,852,262 B2	12/2010	Namineni et al.
7,050,177 B2	5/2006	Tomasi et al.	7,855,732 B2	12/2010	Williams et al.
7,050,606 B2	5/2006	Paul et al.	7,860,301 B2	12/2010	Se et al.
7,058,204 B2	6/2006	Hildreth et al.	RE42,256 E	3/2011	Edwards
7,060,957 B2	6/2006	Lange et al.	7,898,522 B2	3/2011	Hildreth et al.
7,098,435 B2	8/2006	Mueller et al.	7,925,081 B2	4/2011	Gupta et al.
7,113,918 B1	9/2006	Ahmad et al.	7,974,443 B2	7/2011	Kipman et al.
7,121,946 B2	10/2006	Paul et al.	8,009,880 B2	8/2011	Zhang et al.
7,135,992 B2	11/2006	Karlsson et al.	8,031,909 B2	10/2011	Se et al.
7,151,530 B2	12/2006	Roeber et al.	8,035,612 B2	10/2011	Bell et al.
7,158,680 B2	1/2007	Pace	8,035,614 B2	10/2011	Bell et al.
7,162,075 B2	1/2007	Littlefield et al.	8,035,624 B2	10/2011	Bell et al.
7,167,578 B2	1/2007	Blake et al.	8,072,470 B2	12/2011	Marks
7,170,492 B2	1/2007	Bell	8,103,109 B2	1/2012	Winn et al.
7,184,047 B1	2/2007	Crampton	8,144,931 B1	3/2012	Hartman et al.
7,184,048 B2	2/2007	Hunter	8,147,066 B2	4/2012	Nozaki et al.
7,202,898 B1	4/2007	Braun et al.	8,149,268 B1	4/2012	Meyers et al.
7,222,078 B2	5/2007	Abelow	8,154,590 B2	4/2012	Kressel et al.
7,224,384 B1	5/2007	Iddan et al.	8,270,733 B2	9/2012	Cobb et al.
7,227,526 B2	6/2007	Hildreth et al.	8,277,315 B2	10/2012	Burak et al.
7,259,747 B2	8/2007	Bell	8,290,249 B2	10/2012	Mathe et al.
7,293,356 B2	11/2007	Sohn et al.	8,379,919 B2	2/2013	Bronder et al.
7,308,112 B2	12/2007	Fujimura et al.	8,401,225 B2	3/2013	Newcombe et al.
7,310,431 B2	12/2007	Gokturk et al.	8,448,056 B2	5/2013	Pulsipher et al.
7,317,836 B2	1/2008	Fujimura et al.	8,503,720 B2	8/2013	Shotton et al.
7,319,836 B2	1/2008	Kuroda	8,570,320 B2	10/2013	Izadi et al.
7,327,362 B2	2/2008	Grau	8,587,583 B2	11/2013	Newcombe et al.
7,340,077 B2	3/2008	Gokturk et al.	8,711,206 B2	4/2014	Newcombe et al.
7,348,963 B2	3/2008	Bell	9,137,511 B1	9/2015	Legrand et al.
7,359,121 B2	4/2008	French et al.	9,165,199 B2	10/2015	Zhu et al.
7,366,325 B2	4/2008	Fujimura et al.	9,247,238 B2	1/2016	Izadi et al.
7,367,887 B2	5/2008	Watabe et al.	9,256,982 B2	2/2016	Sharp et al.
7,379,563 B2	5/2008	Shamaie	9,262,673 B2	2/2016	Shotton et al.
7,379,566 B2	5/2008	Hildreth	9,940,553 B2	4/2018	Shotton et al.
7,389,591 B2	6/2008	Jaiswal et al.	2002/0024517 A1	2/2002	Yamaguchi et al.
7,412,077 B2	8/2008	Li et al.	2002/0043632 A1	4/2002	Miramonti et al.
7,421,093 B2	9/2008	Hildreth et al.	2002/0069013 A1	6/2002	Navab et al.
7,430,312 B2	9/2008	Gu	2002/0112024 A1	8/2002	Yang et al.
7,436,496 B2	10/2008	Kawahito	2002/0118869 A1	8/2002	Knoploch et al.
7,450,736 B2	11/2008	Yang et al.	2002/0186216 A1	12/2002	Baumberg et al.
7,451,123 B2	11/2008	Platt et al.	2003/0043270 A1	3/2003	Rafey et al.
7,452,275 B2	11/2008	Kuraishi	2003/0067461 A1	4/2003	Fletcher et al.
7,453,472 B2	11/2008	Goede et al.	2003/0161500 A1	8/2003	Blake et al.
7,460,690 B2	12/2008	Cohen et al.	2004/0023612 A1	2/2004	Kriesel
7,489,812 B2	2/2009	Fox et al.	2004/0075738 A1	4/2004	Burke et al.
7,536,032 B2	5/2009	Bell	2004/0104935 A1	6/2004	Williamson et al.
7,555,142 B2	6/2009	Hildreth et al.	2004/0136583 A1	7/2004	Harada et al.
7,560,701 B2	7/2009	Oggier et al.	2004/0145722 A1	7/2004	Uomori et al.
7,570,805 B2	8/2009	Gu	2004/0167667 A1	8/2004	Goncalves et al.
7,574,020 B2	8/2009	Shamaie	2004/0179728 A1	9/2004	Littlefield et al.
7,576,727 B2	8/2009	Bell	2004/0207597 A1	10/2004	Marks
7,583,274 B2	9/2009	Roberts et al.	2004/0233287 A1	11/2004	Schnell
7,590,262 B2	9/2009	Fujimura et al.	2004/0239670 A1	12/2004	Marks
7,593,552 B2	9/2009	Higaki et al.	2005/0010445 A1	1/2005	Krishnan et al.
7,598,942 B2	10/2009	Underkoffler et al.	2005/0059488 A1	3/2005	Larsen et al.
			2005/0078178 A1	4/2005	Brown et al.
			2005/0082480 A1	4/2005	Wagner et al.
			2005/0174473 A1	8/2005	Morgan et al.
			2005/0238200 A1	10/2005	Gupta et al.



(56)

References Cited

U.S. PATENT DOCUMENTS

2006/0006309 A1 1/2006 Dimsdale et al.  
 2006/0018539 A1 1/2006 Sato et al.  
 2006/0064017 A1 3/2006 Krishnan et al.  
 2006/0110021 A1 5/2006 Luo et al.  
 2006/0188144 A1 8/2006 Sasaki et al.  
 2006/0221072 A1 10/2006 Se et al.  
 2006/0221250 A1 10/2006 Rossbach et al.  
 2006/0239553 A1 10/2006 Florin et al.  
 2006/0239558 A1 10/2006 Rafii et al.  
 2006/0252541 A1 11/2006 Zalewski et al.  
 2007/0013718 A1 1/2007 Ohba  
 2007/0016418 A1 1/2007 Mehrotra et al.  
 2007/0031001 A1 2/2007 Hamanaka  
 2007/0052807 A1 3/2007 Zhou et al.  
 2007/0053563 A1 3/2007 Tu et al.  
 2007/0055153 A1 3/2007 Simopoulos et al.  
 2007/0060336 A1 3/2007 Marks et al.  
 2007/0081712 A1 4/2007 Huang et al.  
 2007/0098222 A1 5/2007 Porter et al.  
 2007/0116356 A1 5/2007 Gong et al.  
 2007/0156286 A1 7/2007 Yamauchi  
 2007/0188501 A1 8/2007 Yee et al.  
 2007/0195173 A1 8/2007 Nozaki et al.  
 2007/0216894 A1 9/2007 Garcia et al.  
 2007/0217676 A1 9/2007 Grauman et al.  
 2007/0229498 A1 10/2007 Matusik et al.  
 2007/0260492 A1 11/2007 Feied et al.  
 2007/0260984 A1 11/2007 Marks et al.  
 2007/0276214 A1 11/2007 Dachille et al.  
 2007/0279485 A1 12/2007 Ohba et al.  
 2007/0283296 A1 12/2007 Nilsson  
 2007/0298882 A1 12/2007 Marks et al.  
 2008/0001951 A1 1/2008 Marks et al.  
 2008/0026838 A1 1/2008 Dunstan et al.  
 2008/0027591 A1 1/2008 Lenser et al.  
 2008/0037850 A1 2/2008 Assmann et al.  
 2008/0060854 A1 3/2008 Perlin  
 2008/0062257 A1 3/2008 Corson  
 2008/0075361 A1 3/2008 Winn et al.  
 2008/0075367 A1 3/2008 Winn et al.  
 2008/0100620 A1 5/2008 Nagai et al.  
 2008/0126937 A1 5/2008 Pachet  
 2008/0130985 A1 6/2008 Park et al.  
 2008/0134102 A1 6/2008 Movold et al.  
 2008/0137101 A1 6/2008 Spence et al.  
 2008/0137956 A1 6/2008 Yang et al.  
 2008/0152191 A1 6/2008 Fujimura et al.  
 2008/0152218 A1 6/2008 Okada  
 2008/0215972 A1 9/2008 Zalewski et al.  
 2008/0215973 A1 9/2008 Zalewski et al.  
 2008/0304707 A1 12/2008 Oi et al.  
 2008/0310677 A1 12/2008 Weismuller et al.  
 2008/0310757 A1 12/2008 Wolberg et al.  
 2008/0317331 A1 12/2008 Winn et al.  
 2009/0003652 A1 1/2009 Steinberg et al.  
 2009/0010507 A1 1/2009 Geng  
 2009/0033655 A1 2/2009 Boca et al.  
 2009/0034622 A1 2/2009 Huchet et al.  
 2009/0048482 A1 2/2009 Hong et al.  
 2009/0074238 A1 3/2009 Pfister et al.  
 2009/0096807 A1 4/2009 Silverstein et al.  
 2009/0141933 A1 6/2009 Wagg  
 2009/0167679 A1 7/2009 Klier et al.  
 2009/0175540 A1 7/2009 Dariush et al.  
 2009/0221368 A1 9/2009 Yen et al.  
 2009/0231425 A1 9/2009 Zalewski  
 2009/0244065 A1 10/2009 Storti et al.  
 2009/0244097 A1 10/2009 Estevez  
 2009/0244309 A1 10/2009 Maison et al.  
 2009/0252423 A1 10/2009 Zhu et al.  
 2009/0279756 A1 11/2009 Gindele et al.  
 2009/0290811 A1 11/2009 Imai  
 2010/0045701 A1\* 2/2010 Scott ..... G01S 5/163  
 345/633

2010/0080434 A1 4/2010 Seifert et al.  
 2010/0085352 A1 4/2010 Zhou et al.  
 2010/0085353 A1 4/2010 Zhou et al.  
 2010/0093435 A1 4/2010 Glaser et al.  
 2010/0094460 A1 4/2010 Choi et al.  
 2010/0098328 A1 4/2010 Se et al.  
 2010/0103196 A1 4/2010 Kumar et al.  
 2010/0111370 A1 5/2010 Black et al.  
 2010/0158352 A1 6/2010 Yu et al.  
 2010/0194863 A1 8/2010 Lopes et al.  
 2010/0197390 A1 8/2010 Craig et al.  
 2010/0197393 A1 8/2010 Geiss  
 2010/0197400 A1 8/2010 Geiss  
 2010/0201808 A1 8/2010 Hsu  
 2010/0260396 A1 10/2010 Brandt et al.  
 2010/0278384 A1 11/2010 Shotton et al.  
 2010/0281432 A1 11/2010 Geisner et al.  
 2010/0295783 A1 11/2010 El dokor et al.  
 2010/0296724 A1 11/2010 Chang et al.  
 2010/0302247 A1 12/2010 Perez et al.  
 2010/0302395 A1 12/2010 Mathe et al.  
 2010/0303289 A1 12/2010 Polzin et al.  
 2011/0034244 A1 2/2011 Marks et al.  
 2011/0085705 A1 4/2011 Izadi et al.  
 2011/0096832 A1 4/2011 Zhang et al.  
 2011/0188715 A1 8/2011 Shotton et al.  
 2011/0210915 A1 9/2011 Shotton et al.  
 2011/0228997 A1 9/2011 Sharp et al.  
 2011/0234481 A1 9/2011 Katz et al.  
 2011/0243386 A1 10/2011 Sofka et al.  
 2011/0249865 A1 10/2011 Lee et al.  
 2011/0267344 A1 11/2011 Germann et al.  
 2011/0293180 A1 12/2011 Criminisi et al.  
 2011/0304705 A1 12/2011 Kantor et al.  
 2011/0306398 A1 12/2011 Boch et al.  
 2012/0014559 A1 1/2012 Suehling et al.  
 2012/0056982 A1 3/2012 Katz et al.  
 2012/0062719 A1 3/2012 Debevec et al.  
 2012/0075343 A1 3/2012 Chen et al.  
 2012/0120199 A1 5/2012 Ben Himane  
 2012/0128201 A1 5/2012 Brickhill  
 2012/0147149 A1 6/2012 Liu et al.  
 2012/0147152 A1 6/2012 Vogiatzis et al.  
 2012/0148162 A1 6/2012 Zhang et al.  
 2012/0163656 A1 6/2012 Wang et al.  
 2012/0163723 A1 6/2012 Balan et al.  
 2012/0194516 A1 8/2012 Newcombe et al.  
 2012/0194517 A1 8/2012 Izadi et al.  
 2012/0194644 A1 8/2012 Newcombe et al.  
 2012/0194650 A1 8/2012 Izadi et al.  
 2012/0195471 A1 8/2012 Newcombe et al.  
 2012/0196679 A1 8/2012 Newcombe et al.  
 2012/0212509 A1 8/2012 Benko et al.  
 2012/0239174 A1 9/2012 Shotton et al.  
 2012/0268940 A1 10/2012 Sahlin et al.  
 2012/0300979 A1\* 11/2012 Pirchheim ..... G06T 7/246  
 382/103  
 2012/0314039 A1 12/2012 You et al.  
 2013/0051626 A1 2/2013 Abadpour et al.  
 2013/0077059 A1 3/2013 Marti et al.  
 2013/0243313 A1 9/2013 Civit et al.  
 2013/0251246 A1 9/2013 Tang et al.  
 2013/0265502 A1 10/2013 Huebner  
 2013/0266182 A1 10/2013 Shotton et al.  
 2013/0300740 A1 11/2013 Snyder et al.  
 2013/0316282 A1 11/2013 Ishigami et al.  
 2014/0079314 A1 3/2014 Yakubovich et al.  
 2014/0241617 A1 8/2014 Shotton et al.  
 2015/0029222 A1\* 1/2015 Hofmann ..... G06K 9/00993  
 345/633  
 2016/0163054 A1 6/2016 Izadi et al.  
 2016/0171295 A1 6/2016 Shotton et al.  
 2022/0196840 A1 6/2022 Hilliges et al.

FOREIGN PATENT DOCUMENTS

CN 1738426 A 2/2006  
 CN 1802586 A 7/2006  
 CN 1820509 A 7/2006



(56)

## References Cited

## FOREIGN PATENT DOCUMENTS

CN	101257641	A	9/2008
CN	101305401	A	11/2008
CN	101377812	A	3/2009
CN	201254344	Y	6/2009
CN	101605211	A	12/2009
CN	101254344	B	6/2010
CN	101872491	A	10/2010
CN	102129708	A	7/2011
CN	102184558	A	9/2011
CN	102236912	A	11/2011
CN	102289815	A	12/2011
CN	102609942	A	7/2012
CN	102622762	A	8/2012
CN	102622776	A	8/2012
CN	102681661	A	9/2012
EP	0583061	A2	2/1994
GB	2411532	A	8/2005
GB	2458305	A	9/2009
JP	08044490	A	2/1996
JP	2003141569	A	5/2003
JP	2005210722	A	8/2005
JP	2009113711	A	5/2009
JP	2009536406	A	10/2009
JP	2010079453	A	4/2010
JP	2010279023	A	12/2010
KR	20100131060	A	12/2010
TW	200947346	A	11/2009
WO	9310708	A1	6/1993
WO	9717598	A1	5/1997
WO	9915863	A1	4/1999
WO	9935855	A1	7/1999
WO	9944698	A2	9/1999
WO	0120914	A2	3/2001
WO	0159975	A3	1/2002
WO	02082249	A2	10/2002
WO	03001722	A3	3/2003
WO	03046706	A1	6/2003
WO	03073359	A3	11/2003
WO	03054683	A3	12/2003
WO	03071410	A3	3/2004
WO	2007132451	A2	11/2007
WO	2009059065	A1	5/2009
WO	2009131539	A1	10/2009
WO	2010095080	A1	8/2010
WO	2010140613	A1	12/2010

## OTHER PUBLICATIONS

Wagner, Daniel, et al. "Pose tracking from natural features on mobile phones." Proceedings of the 7th IEEE/ACM International Symposium on Mixed and Augmented Reality. IEEE Computer Society, 2008. (Year: 2008).\*

Wagner, Daniel, Dieter Schmalstieg, and Horst Bischof. "Multiple target detection and tracking with guaranteed framerates on mobile phones." 2009 8th IEEE International Symposium on Mixed and Augmented Reality. IEEE, 2009. (Year: 2009).\*

Chum, Ondřej, Jiří Matas, and Josef Kittler. "Locally optimized RANSAC." Joint Pattern Recognition Symposium. Springer, Berlin, Heidelberg, 2003. (Year: 2003).\*

"Second Office action Issued in Chinese Patent Application No. 201210021583.7", dated Sep. 19, 2014, 8 Pages.

"Third Office Action Issued in Chinese Patent Application No. 201210021583.7", dated Mar. 27, 2015, 8 Pages.

"First Office Action and Search Report Issued in Chinese Patent Application No. 201210431972.7", dated Dec. 1, 2014, 13 Pages.

"Second Office Action Issued in Chinese Patent Application No. 201210431972.7", dated Jul. 7, 2015, 7 Pages.

"Third Office Action Issued in Chinese Patent Application No. 201210431972.7", dated Nov. 26, 2015, 8 Pages.

"First Office Action Issued in Chinese Patent Application No. 201210470054.5", dated Dec. 17, 2014, 17 Pages.

"Second Office Action Issued in Chinese Patent Application No. 201210470054.5", dated Aug. 13, 2015, 6 Pages.

"Office Action Issued in Japanese Patent Application No. 2013-552530", dated Nov. 4, 2015, 5 Pages.

"Office Action Issued in Japanese Patent Application No. 2013-552531", dated Oct. 26, 2015, 6 Pages.

"Office Action Issued in Japanese Patent Application No. 2013-552532", dated Jan. 29, 2016, 6 Pages.

"First Office Action and Search Report Issued in Chinese Patent Application No. 201380068406.7", dated Mar. 2, 2017, 13 Pages.

"Second Office Action Issued in Chinese Patent Application No. 201380068406.7", dated Sep. 20, 2017, 7 Pages.

"Third Office Action Issued in Chinese Patent Application No. 201380068406.7", dated Feb. 27, 2018, 7 Pages.

"First Office Action Issued in Chinese Patent Application No. 201480006047.7", dated Jan. 26, 2017, 14 Pages.

"First Office Action and Search Report Issued in Chinese Patent Application No. 201480010236.1", dated Jan. 19, 2018, 15 Pages.

"Second Office Action Issued in Chinese Patent Application No. 201480010236.1", dated Sep. 5, 2018, 6 Pages.

Agarwal, et al., "3D Human Pose from Silhouettes by Relevance Vector Regression", In Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 2, Jun. 27, 2004, 7 Pages.

Agarwal, et al., "Learning Methods for Recovering 3D Human Pose from Monocular Images", In Technical Report 5333, INRIA Rhone-Aipes, 2004, Oct. 2004, 22 Pages.

Aggarwal, et al., "Human Motion Analysis: A Review", In Proceedings of the IEEE Nonrigid and Articulated Motion Workshop, University of Texas at Austin, Austin, TX., Jun. 16, 1997, 13 Pages.

Alkemade, Remi, "Depth Perception for Augmented Reality using Parallel Mean Shift Segmentation", In Bachelor Thesis, Radboud University., Apr. 27, 2010, 47 Pages.

Amit, et al., "Shape Quantization and Recognition with Randomized Trees", In Journal of Neural Computation, vol. 9, Issue 7, Oct. 1, 1997, 56 Pages.

Anderson, et al., "Augmenting Depth Camera Output Using Photometric Stereo", In Conference on Machine Vision Applications, Jun. 13, 2011, 4 Pages.

Azarbayejani, et al., "Visually Controlled Graphics", In Proceedings of the IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 15, Issue 6, Jun. 1, 1993, 4 Pages.

Baak, et al., "A Data-Driven Approach for Real-Time Full Body Pose Reconstruction from a Depth Camera", In Proceedings with IEEE International Conference on Computer Vision, 2011, Nov. 13, 2011, 8 Pages.

Baatz, et al., "Leveraging 3D City Models for Rotation Invariant Place-of-Interest Recognition", In International Journal of Computer Vision, vol. 96, Issue 3, May 27, 2011, 20 Pages.

Bacon, Pierre-Luc, "Continuous Head Pose Estimation using Random Regression Forests", Retrieved From: <http://pierrelucbacon.com/assets/papers/rrfpose.pdf>, Feb. 9, 2013, 6 Pages.

Ballan, "Marker-less motion capture of skinned models in a four camera set-up using optical flow and silhouettes", 3DPVT, Atlanta, GA, USA, 2008, 8 Pages.

Baltzakis, et al., "Tracking of human hands and faces through probabilistic fusion of multiple visual cues", In Proceedings of 6th International Conference on Computer Vision Systems(ICVS), May 12, 2008, 10 Pages.

Benko, et al., "Depth Touch: Using Depth-Sensing Camera to Enable Freehand Interactions On and Above the Interactive Surface", In Proceedings of IEEE Tabletops and Interactive Surfaces, Oct. 1, 2008, 1 page.

Besl, "A Method for Registration of 3D Shapes", In IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 14, No. 2., Feb. 1992, pp. 239-256.

Beyer, "GPU-based Multi-Volume Rendering of Complex Data in Neuroscience and Neurosurgery", Doctoral Dissertation, Vienna University of Technology., Oct. 2009, pp. 1-117.

Calonder, et al., "BRIEF: Binary Robust Independent Elementary Features", In Proceedings of the 11th European Conference on Computer Vision: Part IV, Sep. 5, 2010, 14 Pages.



(56)

**References Cited**

## OTHER PUBLICATIONS

- Blais, et al., "Registering Multiview Range Data to Create 3D Computer Objects", In Proceedings of IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 17, Issue 8, Aug. 1, 1995, 5 Pages.
- Boehnke, Kay, "Fast Object Localization with Real Time 3D Laser Range Sensor Simulation", In Book World Scientific and Engineering Academy and Society (WSEAS) Transactions On Electronics, Issue 3, vol. 5, Mar. 2008, 10 Pages.
- Bolitho, et al., "Parallel Poisson Surface Reconstruction", In Proceeding of the 5th International Symposium on Advances in Visual Computing (ISVC): Part I, Nov. 26, 2009, 12 Pages.
- Bolitho, Matthew Grant., "The Reconstruction of Large Three-dimensional Meshes", In Dissertation Submitted To The Johns Hopkins University In Conformity With The Requirements For The Degree Of Doctor Of Philosophy, Mar. 2010, 171 Pages.
- Borland, Clarke, "Volumetric Depth Peeling for Medical Image Display", In Visualization and Data Analysis, Jan. 16, 2006, 10 Pages.
- Bosch, et al., "Image Classification using Random Forests and Ferns", In the Proceedings of IEEE 11th International Conference on Computer Vision., Oct. 14, 2007, 8 Pages.
- Botterill, et al., "Bag-of-Words-driven Single Camera SLAM", In Journal of Field Robotics, vol. 28, Issue 2, Oct. 22, 2010, 28 Pages.
- Breen, et al., "Interactive Occlusion and Collision of Real and Virtual Objects in Augmented Reality", In Technical Report of European Computer Industry Research Center-95-02., 1995, 22 Pages.
- Campbell, et al., "Automatic 3D Object Segmentation in Multiple Views using Volumetric Graph-Cuts", In Journal Image and Vision Computing, vol. 28, Issue 1, Jan. 1, 2010, 10 Pages.
- Breiman, et al., "Classification and Regression Trees", Ch. 1, 2, 8, and 11, Monterey, CA: Wadsworth and Brooks, 1984, pp. 1-58, 216-265, and 297-317.
- Breiman, Leo, "Random Forests", In Proceedings of the Machine Learning, vol. 45, Issue 1, Oct. 1, 2001, pp. 5-32.
- Breiman, Leo, "Random Forests—Random Features", In Technical Report 567, University of California, Statistics Department, Berkeley, Sep. 1999, 29 Pages.
- Brogan, et al., "Dynamically Simulated Characters in Virtual Environments", In Proceedings of the IEEE Computer Graphics and Applications, vol. 18, Issue 5, Sep. 1998, 12 Pages.
- Broll, et al., "Toward Next-Gen Mobile AR Games", In Proceedings of IEEE Computer Graphics and Applications, vol. 28, Issue 4, Jul. 9, 2008, 9 Pages.
- Bruckner, Groller, "Exploded Views for Volume Data", In IEEE Transactions on Visualization and Computer Graphics, vol. 12, No. 5, Sep. 12, 2006, 13 Pages.
- Bullitt, Aylward, "Volume Rendering of Segmented Tubular Objects", In Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention, Oct. 14, 2001, pp. 161-168.
- Sequeira, et al., "Automated Reconstruction of 3D Models from Real Environments", Retrieved from: <https://www.sciencedirect.com/science/article/pii/S0924271698000264>, Feb. 1999, 2 Pages.
- Shahar, et al., "Knowledge-Based Visualization of Time-Oriented Clinical Data", In Proceedings of the AMIA Symposium., 1998, 5 Pages.
- Shao, et al., "An Open System Architecture for a Multimedia and Multimodal UserInterface", Japanese Society for Rehabilitation of Persons with Disabilities, Japan, Aug. 24, 1998, 11 Pages.
- Sharp, et al., "Camera pose estimation for 3D reconstruction", U.S. Appl. No. 13/749,497, filed Jan. 24, 2013 (Filed Date), 27 Pages.
- Sharp, Toby, "Implementing Decision Trees and Forests on a GPU", In European Conference on Computer Vision, Lecture Notes in Computer Science, vol. 5305, 2008, pp. 595-608.
- Sheridan, et al., "Virtual Reality Check", In Journal Of Technology Review, vol. 96, Issue 7, Oct. 1993, 9 Pages.
- Shimizu, et al., "Multi-Organ Segmentation in Three-dimensional Abdominal CT Images", In International Journal of Computer Assisted Radiology and Surgery (J CARS), vol. 1., 2006, 8 Pages.
- Shin, et al., "Occlusion Removal Technique for Improved Recognition of Partially Occluded 3D Objects in Computational Integral Imaging", In Journal 3D Research , vol. 1 Issue 2., Jun. 2010, 9 Pages.
- Shivappa, et al., "Person Tracking with Audio-Visual Cues Using Iterative Decoding Framework", In Proceedings of IEEE Fifth International Conference on Advanced Video and Signal Based Surveillance, Sep. 1, 2008, pp. 260-267.
- Shotton, et al., "Real-time Human Pose Recognition in Parts from a Single Depth Image", In Proceedings of the 24th IEEE Conference on Computer Vision and Pattern Recognition, Jun. 2011, 8 Pages.
- Shotton, et al., "Computing pose and/or shape of modifiable entities", U.S. Appl. No. 13/300,542, filed Nov. 18, 2011, 38 Pages.
- Shotton, et al., "Efficient Human Pose Estimation from Single Depth Images", In Proceedings of IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 35, Issue 12, Jun. 21, 2011, 21 Pages.
- Welch, et al., "Motion Tracking: No Silver Bullet, but a Respectable Arsenal", In IEEE Computer Graphics and Application, vol. 22 Issue 6, Nov. 2002, 15 Pages.
- Shotton, et al., "Scene Coordinate Regression Forests for Camera Relocalization in RGB-D Images", In Proceedings with IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Jun. 23, 2013, 8 Pages.
- Shotton, et al., "Semantic Texton Forests for Image Categorization and Segmentation", In proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Jun. 23, 2008, 8 pages.
- Shotton, et al., "TextonBoost for Image Understanding: Multi-Class Object Recognition and Segmentation by Jointly Modeling Texture, Layout, and Context", In International Journal of Computer Vision, vol. 81, Issue 1, Jan. 2009, 30 Pages.
- Silva, et al., "A Survey of GPU-Based vol. Rendering of Unstructured Grids", Revista de Informatica Teorica e Aplicada (RITA), vol. 12, No. 2, 2005, 22 Pages.
- Singh, et al., "Parallel Visualization Algorithms: Performance and Architectural Implications", In IEEE Computer Society Press, Computer, vol. 27, No. 7., Jul. 1994, 11 Pages.
- Smelyanskiy, et al., "Mapping High-Fidelity Volume Rendering for Medical Imaging to CPU, GPU and Many-Core Architectures", In the Proceedings of IEEE Transactions on Visualization and Computer Graphics, vol. 15, No. 6., Nov. 15, 2009, 8 Pages.
- Sminchisescu, et al., "Human Pose Estimation from Silhouettes a Consistent Approach using Distance Level Sets", In Proceedings of WSCG International Conference on Computer Graphics, Visualization and Computer Vision, 2002, 8 Pages.
- Stein, et al., "Structural Indexing: Efficient 3-D Object Recognition", In IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 14, Issue 2., Feb. 1992, 21 Pages.
- Stevens, Jane E., "Flights into Virtual Reality Treating Real World Disorders", In Proceedings of the Washington Post, Science Psychology, Mar. 27, 1995, 2 Pages.
- Strengert, et al., "Large Volume Visualization of Compressed Time-Dependent Datasets on GPU Clusters", Parallel Computing, vol. 31, No. 2, Elsevier Science Publishers, Amsterdam., Feb. 2005, 15 Pages.
- Stuhmer, et al., "Real-Time Dense Geometry from a Handheld Camera", In Journal of Springer-Verlag Berlin, LNCS 6376, Conference on Pattern Recognition (DAGM, Sep. 22, 2010, 10 Pages.
- Sun, et al., "Conditional Regression Forests for Human Pose Estimation", In Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition, Jun. 16, 2012, 8 Pages.
- Tam, et al., "Volume Rendering of Abdominal Aortic Aneurysms", In the Proceedings of the 8th conference on Visualization '97, Phoenix, Arizona, USA., Oct. 24, 1997, 9 Pages.
- Tatarchuk, et al., "Advanced Interactive Medical Visualization on the GPU", In the Journal of Parallel and Distributed Computing, vol. 68 Issue 10, Academic Press, Inc. Orlando, FL, USA., Oct. 2008, 11 Pages.



(56)

**References Cited**

## OTHER PUBLICATIONS

- Taylor, et al., "The Vitruvian Manifold: Inferring Dense Correspondences for One-Shot Human Pose Estimation", In IEEE Conference on Computer Vision and Pattern Recognition, Jun. 16, 2012, 8 Pages.
- Thayananthan, et al., "Pose Estimation and Tracking Using Multivariate Regression", In Journal of Pattern Recognition Letters, vol. 29, Issue 9, Oct. 8, 2007, 15 Pages.
- Thrun, et al., "Probabilistic Robotics", The MIT Press, Chapter 9, Sep. 2005, 56 Pages.
- Tomandl, et al., "Visualization Techniques for Interactive Direct Volume Rendering in Neuroradiology", RadioGraphics, vol. 21, No. 6., Nov. 21, 2007, 12 Pages.
- Torr, et al., "Outlier Detection and Motion Segmentation", In Sensor Fusion VI, vol. 2059, International Society for Optics and Photonics., Aug. 20, 1993, 12 Pages.
- Torralba, "Sharing Visual Features for Multiclass and Multiview Object Detection", In IEEE Transactions on Pattern Analysis and Machine Intelligence., May 29, 2007, 18 Pages.
- Totsuka, et al., "Frequency Domain Volume Rendering", In Proceedings of the 20th Annual Conference on Computer Graphics and Interactive Techniques, Sep. 1, 1993, 8 Pages.
- Tu, et al., "Brain Anatomical Structure Segmentation by Hybrid Discriminative/Generative Models", Retrieved from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2807446/>, Apr. 29, 2008, 24 Pages.
- Tykkala, et al., "Direct Iterative Closest Point for Real-Time Visual Odometry", In Proceedings of the IEEE International Conference on Computer Vision Workshops, Nov. 13, 2011, 7 Pages.
- Vaughan-Nicholas, "Game-Console Makers Battle over Motion-Sensitive Controllers", In IEEE Computer Society, Computer, Aug. 2009, 3 Pages.
- Veas, et al., "Creating Meaningful Environment Models for Augmented Reality", In Proceedings of the IEEE Virtual Reality Conference, Mar. 8, 2008, 2 Pages.
- Vibha, et al., "Classification of Mammograms Using Decision Trees", In Proceedings of IEEE 10th International Database Engineering and Applications Symposium (IDEAS), Computer Society., Dec. 2006., 4 Pages.
- Vidal, et al., "Pursuit-Evasion Games with Unmanned Ground and Aerial Vehicles", In Proceedings of IEEE Inti Conference on Robotics and Automation, May 2001, 8 Pages.
- Viola, et al., "Detecting Pedestrians Using Patterns of Motion and Appearance", In Proceedings of the Ninth IEEE International Conference on Computer Vision, vol. 2., Oct. 13, 2003., 10 Pages.
- Viola, et al., "GPU-based Frequency Domain Volume Rendering", In Proceedings of the ACM 20th spring conference on Computer graphics, 2004, 10 Pages.
- Viola, et al., "Robust Real-Time Face Detection", Published : In International Journal of Computer Vision, vol. 57, Issue 2, May 2004., 18 pages.
- Vitter, et al., "Random Sampling with a Reservoir", In Journal ACM Transactions on Mathematical Software, vol. 11 Issue 1., Mar. 1985, 21 Pages.
- Vogiatzis, et al., "Reconstructing Relief Surfaces", In Image and Vision Computing, vol. 26, Issue 3, Mar. 1, 2008, 10 Pages.
- Waage, et al., "State of the Art Parallel Computing in Visualization using CUDA and OpenCL", The Eurographics Association, Seminar in Visualization., 2008, 7 Pages.
- Wald, "Faster Isosurface Ray Tracing Using Implicit KD-Trees", In Journal IEEE Transactions on Visualization and Computer Graphics archive vol. 11 Issue 5., Sep. 11, 2005, 11 Pages.
- Wang, et al., "2D Face Fitting-assisted 3D face reconstruction for pose-robust face recognition", In Soft Computing, A Fusion of Foundations, Methodologies and Applications, Springer. Berlin, vol. 15, No. 3, Nov. 8, 2009, pp. 417-428.
- Wang, et al., "Viewpoint Invariant 3D Landmark Model Inference from Monocular 2D Images Using Higher-Order Priors", In the Proceedings of the 2011 International Conference on Computer Vision, IEEE Computer Society Washington, DC, USA, Nov. 6-13, 2011, pp. 319-326.
- "A Framework for Intelligent Visualization of Multiple Time-Oriented Medical Records", In the Proceedings of Annual Symposium Proceedings, AMIA, vol. 2005., 2005, 9 Pages.
- "Channel Access Method", Retrived From: [http://en.wikipedia.org/wiki/Multiple\\_access\\_protocol#Circuit\\_mode\\_and\\_channelization\\_methods.](http://en.wikipedia.org/wiki/Multiple_access_protocol#Circuit_mode_and_channelization_methods.), Oct. 26, 2011, 7 Pages.
- "Data Explosion: The Challenge of Multidetector-Row CT", In European Journal of Radiology, vol. 36, Issue 2., Nov. 2000, pp. 74-80.
- "fovia.com", Retrieved from: <https://www.fovia.com/>, 2005, 3 Pages.
- "From Structure-from-Motion Point Clouds to Fast Location Recognition", In Proceeding of IEEE Conference on Computer Vision and Pattern Recognition, Jun. 20, 2009, 8 Pages.
- "Kabsch algorithm", Retrieved from: [en.wikipedia.org/wiki/kabsch\\_algorithm](http://en.wikipedia.org/wiki/kabsch_algorithm), May 16, 2013 (Retrieved Date), 3 Pages.
- "Manipulator and Object Tracking for In Hand Model Acquisition", In Proceedings, IEEE International Conference on Robots and Automation, May 7, 2010, 8 Pages.
- "MapReduce", Retrieved from :<http://web.archive.org/web/20120111070052/http://en.wikipedia.org/wiki/MapReduce>, Jan. 11, 2012 (Retrived Date), 5 Pages.
- "Signed distance function", Retrieved from :[http://web.archive.org/web/20120120095205/http://en.wikipedia.org/wiki/Signed\\_distance\\_function](http://web.archive.org/web/20120120095205/http://en.wikipedia.org/wiki/Signed_distance_function), Jan. 20, 2012 (Retrieved Date), 1 Page.
- "Simulation and Training", Division Incorporated, Dec. 1994, 6 Pages.
- "ThevisualMD", Retrieved from: <http://www.thevisualmd.com/>, retrieved on Jan. 15, 2010, 1 Page.
- "Non-Final Office Action Issued in U.S. Appl. No. 13/017,474", dated Feb. 5, 2013, 14 Pages.
- "Non-Final Office Action Issued in U.S. Appl. No. 13/017,518", dated Sep. 25, 2013, 11 Pages.
- Criminisi, et al., "Volume Rendering on Server GPUs for Enterprise-Scale Medical Applications", In Proceedings of IEEE-VGTC Symposium on Visualization 2010, vol. 29 (2010), No. 3., 2010, 10 Pages.
- "Office Action Issued in Taiwan Patent Application No. 101101731", dated Apr. 24, 2012, 15 Pages.
- "Final Office Action Issued in U.S. Appl. No. 13/017,518", dated May 20, 2014, 14 Pages.
- "Office Action in Korean Patent Application No. 10-2013-7019975", dated Feb. 12, 2018, 4 Pages.
- "Office Action issued in Korean Patent Application No. 10-2013-7020295", dated Aug. 4, 2017, 12 Pages.
- "Final Office Action Issued in U.S. Appl. No. 12/454,628", dated Oct. 5, 2012, 9 Pages.
- "Non Final Office Action Issued in U.S. Appl. No. 12/454,628", dated May 4, 2012, 9 Pages.
- "Final Office Action Issued in U.S. Appl. No. 12/697,785", dated Mar. 29, 2013, 13 Pages.
- "Final Office Action Issued in U.S. Appl. No. 12/697,785", dated Aug. 20, 2014, 14 Pages.
- "Non-Final Office Action Issued in U.S. Appl. No. 12/697,785", dated Mar. 25, 2014, 14 Pages.
- "Non-Final Office Action Issued in U.S. Appl. No. 12/697,785", dated Sep. 13, 2012, 12 Pages.
- "Final Office Action Issued in U.S. Appl. No. 12/725,811", dated Apr. 30, 2015, 20 Pages.
- "Final Office Action Issued in U.S. Appl. No. 12/725,811", dated Sep. 11, 2013, 18 Pages.
- "Non-Final Office Action Issued in U.S. Appl. No. 12/725,811", dated Feb. 25, 2013, 17 Pages.
- "Non-Final Office Action Issued in U.S. Appl. No. 12/725,811", dated Jul. 15, 2014, 20 Pages.
- "Final Office Action Issued in U.S. Appl. No. 12/979,362", dated Aug. 29, 2013, 12 Pages.
- "Final Office Action Issued in U.S. Appl. No. 12/979,362", dated Sep. 25, 2014, 12 Pages.
- "Non-Final Office Action Issued in U.S. Appl. No. 12/979,362", dated Mar. 14, 2014, 11 Pages.
- "Non-Final Office Action Issued in U.S. Appl. No. 12/979,362", dated Apr. 9, 2015, 13 Pages.



(56)

## References Cited

## OTHER PUBLICATIONS

- “Non-Final Office Action Issued in U.S. Appl. No. 12/979,362”, dated Mar. 19, 2013, 12 Pages.
- “Office Action Issued in European Patent Application No. 12190592.1”, dated Apr. 10, 2013, 5 Pages.
- “Search Report Issued in European Patent Application No. 12190592.1”, dated Mar. 28, 2013, 3 Pages.
- “Office Action Issued in European Patent Application No. 12191941.9”, dated Jan. 3, 2018, 9 Pages.
- “Search Report Issued in European Patent Application No. 12191941.9”, dated Dec. 7, 2017, 9 Pages.
- “Search Report Issued in European Patent Application No. 12191941.9”, dated Aug. 1, 2017, 15 Pages.
- “Extended European Search Report Issued in European Patent Application No. 12741545.3”, dated Jun. 27, 2014, 3 Pages.
- “Office Action Issued in European Patent Application No. 12741545.3”, dated Jul. 14, 2014, 4 Pages.
- “Office Action Issued in European Patent Application No. 12741545.3”, dated Feb. 9, 2018, 4 Pages.
- “European Search Report issued in European Patent Application No. 12741552.9”, dated Jan. 3, 2017, 4 Pages.
- “Office Action Issued in European Patent Application No. 12741552.9”, dated Mar. 17, 2017, 7 Pages.
- “Office Action Issued in European Patent Application No. 12741552.9”, dated Apr. 26, 2018, 5 Pages.
- “Office Action Issued in European Patent Application No. 12742358.0”, dated Feb. 2, 2017, 7 Pages.
- “Search Report Issued in European Patent Application No. 12742358.0”, dated Jan. 9, 2017, 4 Pages.
- “Final Office Action Issued in U.S. Appl. No. 13/017,474”, dated Aug. 1, 2013, 15 Pages.
- Kohler, Markus, “Technical Details and Ergonomical Aspects of Gesture Recognition Applied in Intelligent Home Environments”, Retrieved from Internet: <https://pdfs.semanticscholar.org/0210/daa859d8574b1e1f98f809814e83022c75e4.pdf>, Jan. 1997, 35 Pages.
- Kohler, Markus, “Vision Based Remote Control in Intelligent Home Environments”, In Journal of 3D Image Analysis and Synthesis, vol. 96, Nov. 1996, 8 Pages.
- Krahnstoeber, et al., “Articulated Models from Video”, In the Proceedings of the 2004 IEEE Computer Society Conference, vol. 1, pp. I-I., Jun. 2004, 8 Pages.
- Krainin, et al. “Manipulator and Object Tracking for In Hand Model Acquisition”, In Journal of ICRA 2010 Workshop Paper, 2010, 34 Pages.
- Kruger, et al. “Acceleration Techniques for GPU-based Volume Rendering”, In the Proceeding VIS '03 Proceedings of the 14th IEEE Visualization 2003 (VIS'03), Oct. 22-24, 2003, 6 Pages.
- Kurihara, “Modeling Deformable Human Hands from Medical Images”, Proceedings of the 2004 ACM SIGGRAPH, 2004, 9 Pages.
- Kurkure, et al., “Automated Segmentation of Thoracic Aorta in Non-Contrast CT Images”, IEEE International Symposium on Biomedical, May 14, 2008, 4 Pages.
- Lacroute, et al., “Fast Volume Rendering Using a Shear-Warp Factorization of the Viewing Transformation”, In the Proceedings of the 21st Annual Conference on Computer Graphics and Interactive Techniques, Jul. 24, 1994, 236 Pages.
- Lacroute, “Real-Time Volume Rendering on Shared Memory Multiprocessors Using the Shear-Warp Factorization”, In the Proceedings of the IEEE Symposium on Parallel Rendering, Atlanta, Georgia, USA., Oct. 30-31, 1995, 6 Pages.
- Lai, et al., “Sparse Distance Learning for Object Recognition Combining RGB and Depth Information”, Retrieved from: <http://ils.intel-research.net/publications/47>, 2011, 7 Pages.
- Laxton, Benjamin, “Monocular Human Pose Estimation”, Retrieved From <<[http://vision.ucsd.edu/~blaxton/pagePapers/laxton\\_researchExam2007.pdf](http://vision.ucsd.edu/~blaxton/pagePapers/laxton_researchExam2007.pdf)>>, Jun. 26, 2009, 16 Pages.
- Lay, et al., “Process Oriented Knowledge Management to Support Clinical Pathway Execution”, In Wissens Management, Apr. 22, 2005.
- Le Grand “Broad-Phase Collision Detection with CUDA”, In journal of GPU Gems 3, Chapter 32, 2008, 28 Pages.
- Lee, et al. “Body Part Detection for Human Pose Estimation and Tracking”, In Proceedings of the IEEE Workshop on Motion and Video Computing, Feb. 23, 2007, 8 Pages.
- Leibe, et al. “Robust Object Detection with Interleaved Categorization and Segmentation”, In International Journal of Computer Vision, vol. 77 Issue 1-3., May 2008, 31 Pages.
- Lepetit, et al. “Keypoint Recognition using Randomized Trees”, In IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 28, Issue 9., Sep. 2006, 15 Pages.
- Levoy, et al. “Display of Surfaces from Volume Data”, In Journal IEEE Computer Graphics and Applications Archive vol. 8 Issue 3., May 1988, 18 Pages.
- Levoy, et al. “The Digital Michelangelo Project: 3D Scanning of Large Statues”, In Proceedings of ACM SIGGRAPH, Jul. 2000, 14 Pages.
- Levoy, et al. “Volume Rendering by Adaptive Refinement”, In Journal The Visual Computer: International Journal of Computer Graphics archive vol. 6 Issue 1., 1990, 5 Pages.
- Li, et al. “Empty Space Skipping and Occlusion Clipping for Texture-based Volume Rendering”, In the Proceedings of the 14th IEEE Visualization 2003 (VIS'03), IEEE Computer Society Washington, DC, USA., Oct. 22-24, 2003, 8 Pages.
- Li, et al. “Location Recognition using Prioritized Feature Matching”, In Proceedings of the 11th European Conference on Computer Vision, Sep. 5, 2010, 14 Pages.
- Linguraru, “Multi-organ Automatic Segmentation in 4D Contrast-Enhanced Abdominal CT”, In IEEE International Symposium on Biomedical Imaging From Nano to Macro, May 14, 2008, 4 Pages.
- Liu, Ting, et al. “An Investigation of Practical Approximate Nearest Neighbor Algorithms”, Advances in neural information processing systems, 2004, 8 Pages.
- Livingston, Mark Alan. “Vision-based Tracking with Dynamic Structured Light for Video See-through Augmented Reality”, A Dissertation Submitted to the Faculty of The University of North Carolina at Chapel Hill in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy in the Department of Computer Science, Jan. 1998, 145 Pages.
- Ljung, et al. “Efficient Methods for Direct Volume Rendering of Large Data Sets”, Doctoral Thesis, Linköping University, Department of Science and Technology, Visual Information Technology and Applications (VITA). Linköping University, The Institute of Technology., 2006, 90 Pages.
- Lorensen, et al. “Marching Cubes: A High Resolution 3d Surface Construction Algorithm”, In Proceedings of 14th Annual Conference on Computer Graphics and Interactive Techniques, vol. 21, Issue 4, Aug. 1, 1987, 7 Pages.
- Lu, et al. “Multi-view human motion capture with an improved deformation skin model”, Digital Image Computing: Techniques and Applications, 2008, 8 Pages.
- Lum, et al. “Texture Hardware Assisted Rendering of Time-Varying Volume Data”, In the Proceedings of the Conference on Visualization '01, IEEE Computer Society Washington, DC, USA., Oct. 21-26, 2001, 10 Pages.
- Lysenkov, et al. “Recognition and Pose Estimation of Rigid Transparent Objects with a Kinect Sensor”, In Robotics: Science and Systems 2012, Sydney, NSW, Australia., Jul. 13, 2012, 8 Pages.
- Ma, et al. “A Data Distributed, Parallel Algorithm for Ray-Traced Volume Rendering”, In the Proceedings of the 1993 Symposium on Parallel Rendering, San Jose, California, USA., Oct. 25-26, 1993, 32 Pages.
- Magenat-Thalmann, et al. “Joint-dependent local deformations for hand animation and object grasping”, In Proceedings on Graphics interface, Canadian Information Processing Society, 1988, 12 Pages.
- Mair, et al. “Efficient camera-based pose estimation for real-time applications”, In International Conference on Intelligent Robots and Systems, Oct. 10, 2009, 8 Pages.
- Markelj, et al. “A Review of 3D/2D Registration Methods for Image-Guided Interventions”, Medical Image Analysis, Oxford University Press, Oxford, GB, vol. 16, No. 3., Apr. 1, 2012, pp. 642-661.



(56)

## References Cited

## OTHER PUBLICATIONS

- Michel, et al. "GPU-Accelerated Real-time 3D Tracking for Humanoid Locomotion and Stair Climbing", Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, Nov. 2007,,7 Pages.
- Milella, et al. "Stereo-Based Ego-Motion Estimation Using Pixel Tracking and Iterative Closest Point", In Proceedings of the Fourth IEEE International Conference on Computer Vision Systems, Jan. 4, 2006,7 Pages.
- Miyagawa, et al. "CCD-Based Range Finding Sensor", In IEEE Transactions on Electron Devices, vol. 44 , Issue: 10, Oct. 1997,pp. 1648-1652.
- Molchanov, et al. "Non-iterative Second-order Approximation of Signed Distance Functions for Any Isosurface Representation", In Eurographics/ IEEE-VGTC Symposium on Visualization, vol. 29, Issue 3, Nov. 2010,10 Pages.
- Montillo, et al. "Age Regression from Faces Using Random Forests", In Proceedings of the 16th IEEE International Conference on Image Processing, Nov. 7, 2009,4 Pages.
- Moreno, et al. "A Contextual Medical Image Viewer", In the Proceedings of IEEE Transactions of Information Technology in Biomedicine, vol. 11, No. 5, Sep. 2007,10 Pages.
- Mori "Guiding Model Search Using Segmentation", In the Proceedings of the IEEE Workshop on Motion and Video Computing, IEEE Computer Society Washington, DC, USA., Feb. 23-24, 2007,7 Pages.
- Morra, et al. "Automatic Subcortical Segmentation Using a Contextual Model", In Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention, Sep. 6, 2008,8 Pages.
- Narayanan, et al. "Ray Casting Deformable Models on the GPU", In Sixth Indian Conference on Computer Vision, Graphics & Image Processing., Dec. 16, 2008,9 Pages.
- Navaratnam, et al. "Hierarchical Part-Based Human Body Pose Estimation", In Proceedings of British Machine Vision Association (BMVC), Sep. 5, 2005,10 Pages.
- Neumann, et al. "Interactive Volume Rendering on a Multi-Computer", In Proceedings of the 1992 symposium on Interactive 3D Graphics, Jun. 1, 1992,8 Pages.
- Newcombe, et al. "DTAM: Dense Tracking and Mapping in Real-Time", In Proceedings of IEEE International Conference on Computer Vision, Nov. 6, 2011,8 Pages.
- Newcombe, et al. "KinectFusion: Real-Time Dense Surface Mapping and Tracking", In Proceedings of the 10th IEEE International Symposium on Mixed and Augmented Reality, Oct. 26, 2011,10 Pages.
- Newcombe, et al. "Live Dense Reconstruction with a Single Moving Camera", In Proceedings of The Twenty-Third IEEE Conference on Computer Vision and Pattern Recognition, Jun. 13, 2010,8 Pages.
- Newcombe, et al. "Mobile Camers Localization using Depth Maps.", Retrieved from: <https://patents.google.com/patent/US8711206B2/en>, Apr. 29, 2014,25 Pages.
- Newcombe, et al. "Moving Object Segmentation Using Depth Images.", Retrieved from: <https://patents.google.com/patent/US8401225B2/en>, Mar. 19, 2013,32 Pages.
- Newcombe, et al. "Real-Time Camera Tracking using Depth Maps", Retrieved from: <https://patents.google.com/patent/US8401242B2/en>, Mar. 19, 2013,33 Pages.
- Gemeiner, et al. "Improving Localization Robustness in Monocular SLAM Using a High-Speed Camera", In Proceedings of Robotics: Science and Systems, Jun. 25, 2008,8 Pages.
- Gemme, et al. "Pose Refinement Using ICP Applied to 3-D LIDAR Data for Exploration Rovers", In Proceedings of International Symposium on Artificial Intelligence, Robotics and Automation in Space, Sep. 4, 2012,8 Pages.
- Ginneken "Robust Segmentation and Anatomical Labeling of the Airway Tree from Thoracic CT Scans", In International Conference on Medical Image Computing and Computer-Assisted Intervention, Sep. 6, 2008,pp. 219-226.
- Goesele, et al. "Multiview Stereo Revisited", In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2006,8 Pages.
- Granieri, et al. "Simulating Humans in VR", Published in the British Computer Society, Academic Press, Oct. 12, 1994,15 Pages.
- Guthe, et al. "Interactive Rendering of Large Volume Data Sets", In the Proceedings of the Conference on Visualization '02, Boston, Massachusetts., Oct. 27-Nov. 1, 2002,pp. 53-60.
- Hadwiger, et al. "Advanced Illumination Techniques for GPU-Based Volume Raycasting", In International Conference on Computer Graphics and Interactive Techniques., Aug. 2009,56 Pages.
- Hadwiger, et al. "GPU-Based Volume Ray-Casting with Advanced Illumination", In Proceedings of ACM SIGGRAPH 2009 Courses, Article No. 2 , New Orleans, Louisiana., Aug. 3-7, 2009,173 Pages.
- Han, et al. "Atlas-Based Auto-Segmentation of Head and Neck CT Images", In Proceedings of the 11th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), New York, NY., Sep. 6, 2008,8 Pages.
- Haqiqat et, al., "Using Image Moments for Tracking Rotating Objects", In ICGST International Conference on Automation, Robotics and Autonomous Systems, Dec. 19, 2005, pp. 38-43.
- Harada, "Real-Time Rigid Body Simulation on GPUs", In Journal of GPU Gems 3, Chapter 29, 2008, 2007,25 Pages.
- Hardenbergh, et al. "Integrated 3D Visualization of fMRI and DTI Tractography", In Proceedings of IEEE Visualization, Oct. 2005,2 Pages.
- Hasegawa, et al. "Human-Scale Haptic Interaction with a Reactive Virtual Human in a Real-Time Physics Simulator", In Proceedings of the ACM Computers in Entertainment, vol. 4, Issue 3, Article 6C, Jul. 1, 2006,12 Pages.
- Hastreiter, et al. "Interactive Direct Volume Rendering of the Inner Ear for the Planning of Neurosurgery", In Bildverarbeitung für die Medizin, Springer, Berlin, Heidelberg, 1999,pp. 192-196.
- He, Lei "Generation of Human Body Models", In Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of Master of Science in Computer Science, University of Auckland, New Zealand, Apr. 2005,111 Pages.
- Kohler, Markus "Special Topics of Gesture Recognition Applied in Intelligent Home Environments", In Proceedings of the International Gesture Workshop on Gesture and Sign Language in Human-Computer Interaction, Sep. 17, 1997,12 Pages.
- Henry, et al. "RGB-D Mapping: Using Depth Cameras for Dense 3D Modeling of Indoor Environments", In the 12th International Symposium on Experimental Robotics, Dec. 18-21, 2010,15 Pages.
- Henry,, et al. "RGB-D mapping: Using Kinect-style Depth Cameras for Dense 3D Modeling of Indoor Environments", In International Journal of Robotic Research, vol. 31, Issue 5., Apr. 2012,28 Pages.
- Herath, et al. "Simultaneous Localisation and Mapping: A Stereo Vision Based Approach", In International Conference on Intelligent Robots and Systems, Oct. 2006,,6 Pages.
- Hersh, et al. "Medical Image Retrieval and Automatic Annotation: OHSU at ImageCLEF 2007", In Workshop of the Cross-Language Evaluation Forum for European Languages, 2007, Sep. 19, 2007,8 Pages.
- Higo, et al. "A Hand-held Photometric Stereo Camera for 3-D Modeling", In Proceedings of IEEE 12th International Conference on Computer Vision, Sep. 29, 2009,8 Pages.
- Hile, et al. "Information Overlay for Camera Phones in Indoor Environments", In Proceedings of the 3rd International Conference on Location—and Context—Awareness, Sep. 20, 2007,17 Pages.
- Hirsch, et al. "BiOi Screen: A Thin, Depth-Sensing LCD for 3D Interaction using Light Fields", In ACM Transactions on Graphics, Proceedings of ACM SIGGRAPH Asia, vol. 28, Issue 5, Dec. 2009,9 Pages.
- Hogue, et al. "Underwater Environment Reconstruction using Stereo and Inertial Data", In IEEE International Conference on Systems, Man and Cybernetics, Jan. 2008,6 Pages.
- Holzer, et al. "Learning to Efficiently Detect Repeatable Interest Points in Depth Data", In Proceedings of the 12th European Conference on Computer Vision, vol. Part I, Oct. 7, 2012,14 Pages.
- Homann, et al. "Vasculature Segmentation of CT Liver Images Using Graph-cuts and Graph-based Analysis", In Proceedings of



(56)

## References Cited

## OTHER PUBLICATIONS

IEEE International Symposium on Biomedical Imaging From Nano to Macro (ISBI), May 14, 2008, 4 Pages.

Hongo, et al. "Focus of Attention for Face and Hand Gesture Recognition Using Multiple Cameras", In Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition, Mar. 28, 2000, 6 Pages.

Huang, et al. "Estimating Human Pose from Occluded Images", In Asian Conference on Computer Vision, Springer, Berlin, Heidelberg, Sep. 23, 2009, 12 Pages.

Humphreys, et al. "Chromium: A Stream Processing Framework for Interactive Rendering on Clusters", In the Proceedings of the 29th Annual Conference on Computer Graphics and Interactive Techniques, San Antonio, Texas, Jul. 23-26, 2002, pp. 693-702.

Isard, et al. "Condensation—Conditional Density Propagation for Visual Tracking", In Proceedings of the International Journal of Computer Vision, vol. 29, Issue 01, Aug. 1998, 24 Pages.

Izadi, et al. "KinectFusion: Real-time 3D Reconstruction and Interaction Using a Moving Depth Camera", In Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology, Oct. 16, 2011, 10 Pages.

Izadi, et al. "Reducing Interference Between Multiple Infra-Red Depth Cameras", Retrieved from: <https://patents.google.com/patent/US9247238B2/en>, Jan. 26, 2016.

Izadi "Using a Three Dimensional Environment Model in Gameplay", Retrieved from: <https://patents.google.com/patent/US8570320B2/en>, Oct. 29, 2013, 31 Pages.

Jakobovits, et al. "WIRM: An Open Source Toolkit for Building Biomedical Web Applications", In Journal of the American Medical Informatics Association, vol. 9, Issue 6, Nov. 2002, 14 Pages.

Jin, et al. "Active illumination based 3D surface reconstruction and registration for image-guided medialization laryngoplasty", In the Proceedings of Optical Diagnostics of Living Cells II, vol. 6509, Mar. 6, 2007, 13 Pages.

Jivet, et al. "Real Time Representation of 3D Sensor Depth Images", WSEAS Transactions on Electronics, vol. 5, Issue 3, Mar. 2008, 7 Pages.

Jung "A Framework of Context-Sensitive Visualization for User-Centered Interactive Systems", In the Proceedings of 10th International Conference on User Modelling, Edinburgh, UK, Jul. 24, 2005, 5 Pages.

Kabsch, W. "A Solution for the Best Rotation to Relate Two Sets of Vectors", Acta Crystallographica Section A: Crystal Physics, Diffraction, Theoretical and General Crystallography, vol. 32, Issue 5, Sep. 1976, pp. 922-923.

Kainz, et al. "Ray Casting of Multiple Volumetric Datasets with Polyhedral Boundaries on Manycore GPUs", In the Proceeding ACM SIGGRAPH Asia 2009 papers Article No. 1, Yokohama, Japan, Dec. 16-19, 2009.

Kalra, et al. "Topological Modeling Of Human Anatomy Using Medical Data", In Proceedings of the Computer Animation Conference, Geneva, Switzerland, Apr. 19, 1995, 12 Pages.

Kanade, et al. "A Stereo Machine for Video-rate Dense Depth Mapping and Its New Applications", In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1996, pp. 196-202.

Kanaujia, et al. "Semi-supervised Hierarchical Models for 3D Human Pose Reconstruction", In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2007, Jun. 17, 2007, 8 Pages.

Kazhdan, M. "Poisson Surface Reconstruction", In Proceedings of the fourth Eurographics symposium on Geometry processing. 2006, Jun. 2006, 10 Pages.

Kil, et al. "GPU-assisted Surface Reconstruction on Locally-Uniform Samples", In Proceedings of the 17th International Meshing Roundtable, Oct. 12, 2008, 18 Pages.

Kim, et al. "Relocalization Using Virtual Keyframes For Online Environment Map Construction", In Proceedings of the 16th ACM Symposium on Virtual Reality Software and Technology, Nov. 2009, 8 Pages.

Klein, et al. "Improving the Agility of Keyframe-Based SLAM", In Proceedings of the 10th European Conference on Computer Vision, Oct. 12, 2008, 14 Pages.

Klein, et al. "Parallel Tracking and Mapping for Small AR Workspaces", In IEEE, International Symposium on Mixed and Augmented Reality, Nov. 2007, 10 Pages.

Kniss, et al. "Multi-Dimensional Transfer Functions for Interactive Volume Rendering", In Journal IEEE Transactions on Visualization and Computer Graphics archive vol. 8 Issue 3, Jul. 2002, pp. 270-285.

Knoop, et al. "Sensor Fusion for 3D Human Body Tracking with an Articulated 3D Body Model", In Proceedings of IEEE International Conference on Robotics and Automation (ICRA), May 19, 2006, 7 Pages.

Carmody, Tim, "How Motion Detection Works in Xbox Kinect", Retrieved from Internet: <http://gizmodo.com/5681078/how-motion-detection-works-in-xbox-kinect>, Nov. 3, 2010, 4 Pages.

Castle, et al., "Towards Simultaneous Recognition, Localization and Mapping for Hand-Held and Wearable Cameras", In Proceedings 2007 IEEE International Conference on Robotics and Automation, Apr. 10, 2007, 6 Pages.

Castle, et al., "Video-Rate Localization in Multiple Maps for Wearable Augmented Reality", In 12th IEEE International Symposium on Wearable Computers, Sep. 28, 2008, 8 Pages.

Celix, "Monocular Vision SLAM for Indoor Aerial Vehicles", In Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, Oct. 11, 2009, 8 Pages.

Chen, et al., "Model Based Object Recognition by Robust Information Fusion", In Proceedings of the 17th International Conference on Pattern Recognition, vol. 3, Aug. 23-26, 2004, 4 Pages.

Chen, et al., "Object Modeling by Registration of Multiple Range Images", In Journal Image and Vision Computing—Special issue: range image understanding archive, vol. 10, Issue 3, Apr. 1, 1992, 6 Pages.

Cheung, et al., "Robust Background Subtraction with Foreground Validation for Urban Traffic Video", In Journal EURASIP Journal on Applied Signal Processing, vol. 2005, Jan. 1, 2005, 12 Pages.

Chiu, et al., "Improving the Kinect by Cross-Modal Stereo", In Proceedings of 22nd British Machine Vision Conference, Aug. 2011, 10 Pages.

Chum, et al., "Locally Optimized RANSAC", In Proceeding of 25th DAGM Symposium, Sep. 10, 2003, 8 Pages.

Cohen, et al., "Interactive Fluid-Particle Simulation using Translating Eulerian Grids", In Proceedings of the 2010 ACM SIGGRAPH symposium on Interactive 3D Graphics and Games, Feb. 19, 2010, 8 Pages.

Comaniciu, et al., "Mean Shift: A Robust Approach Toward Feature Space Analysis", In IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, Issue 5, May 2002, 17 Pages.

Criminisi, et al., "Automatic Semantic Parsing of CT Scans via Multiple Randomized Decision Trees", In Radiological Society of North America (RSNA), PowerPoint presentation, Dec. 2009, 2 Pages.

Criminisi, et al., "Decision Forests with Long-Range Spatial Context for Organ Localization in CT Volumes", In Medical Image Computing and Computer-Assisted Intervention (MICCAI), Sep. 20, 2009, 319 Pages.

Criminisi, et al., "Decision Forests: A Unified Framework for Classification, Regression, Density Estimation, Manifold Learning and Semi-Supervised Learning", In Journal of Foundations and Trends in Computer Graphics and Vision, vol. 7, Issue 2-3, Feb. 2012, 150 Pages.

Criminisi, et al., "GeoS: Geodesic Image Segmentation", In Proceedings of the 10th European Conference on Computer Vision, Part I, Oct. 20, 2008, 14 Pages.

Criminisi, et al., "Regression Forests for Efficient Anatomy Detection and Localization in CT Studies", In Proceedings of the 2010 International MICCAI Conference on Medical Computer Vision: Recognition Techniques and Applications in Medical Imaging, Sep. 20, 2010, pp. 106-117.

Crow, "Summed-Area Tables for Texture Mapping", In Proceeding of SIGGRAPH '84 11th Annual Conference on Computer Graphics and Interactive Techniques, vol. 18 Issue 3, Jul. 1984, 6 Pages.



(56)

**References Cited**

## OTHER PUBLICATIONS

Curless, et al., "A Volumetric Method for Building Complex Models from Range Images", In Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques, Aug. 1, 1996, 10 Pages.

Cutts, Matt, "Matt Cutts: Gadgets, Google, and SEO", Retrieved from Internet:<https://web.archive.org/web/20101129163015/http://www.mattcutts.com/blog/>, Nov. 18, 2010, 10 Pages.

Dalal, et al., "Histograms of oriented gradients for human detection", In Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Jun. 20, 2005, 8 Pages.

Dam, et al., "Immersive VR for Scientific Visualization: A Progress Report", In Journal IEEE Computer Graphics and Applications, vol. 20 Issue 6, Nov. 1, 2000, 27 Pages.

Daniele, Ravi, "Kinect: The Next Generation of Motion Control", Retrieved from:<http://www.dmi.unict.it/~battiato/CVision1112/Kinect.pdf>, Feb. 9, 2013, 66 Pages.

Davison, et al., "Mobile Robot Localisation using Active Vision", In Proceedings of 5th European Conference on Computer Vision, vol. II, Jun. 1998, 17 Pages.

Debevec, et al., "Acquiring the Reflectance Field of a Human Face", In Proceedings of the 27th annual conference on Computer graphics and interactive techniques, Apr. 4, 2004., 12 Pages.

Angel, et al., "Intuitive Volume Classification in Medical Augmented Reality (AR)", Current Topics in Computer and Robot Assisted Surgery (CURAC), German Medical Science., Sep. 20, 2006, 9 Pages.

Ding, et al., "Extraction of Human Body Skeleton Based on Silhouette Images", In Second International Workshop on Education Technology and Computer Science, vol. 1, Mar. 6, 2010, pp. 71-74.

Dolejsi, et al., "Reducing False Positive Responses in Lung Nodule Detector System by Asymmetric Ada boost", In IEEE International Symposium on Biomedical Imaging From Nano to Macro (ISBI), May 14, 2008, 4 Pages.

Dong, et al., "Keyframe-Based Real-Time Camera Tracking", In IEEE 12th International Conference on Computer Vision, Sep. 29, 2009, 8 Pages.

Eade, et al., "Unified Loop Closing and Recovery for Real Time Monocular SLAM", In Proceeding of 19th British Conference on Machine Vision, Sep. 1, 2008, 10 Pages.

Einarsson, et al., "Relighting Human Locomotion with Flowed Reflectance Fields", In the Proceedings of Eurographics Symposium on Rendering, 2006, 2 Pages.

Elfes, et al., "Sensor Integration for Robot Navigation: Combining Sonar and Stereo Range Data in a Grid-Based Representation", In Proceedings of the 26th Conference on Decision and Control, Dec. 1987, 8 Pages.

Elgammal, "Inferring 3D Body Pose from Silhouettes using Activity Manifold Learning", In Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 2, Jun. 27, 2004, 8 Pages.

Engel, et al., "A Framework for Interactive Hardware Accelerated Remote 3D-Visualization", In Data Visualization 2000., May 2000, pp. 167-177.

Engel, et al., "Combining Local and Remote Visualization Techniques for Interactive Volume Rendering in Medical Applications", In the Proceedings of the Conference on Visualization'00. IEEE Computer Society Press., Oct. 1, 2000, 5 Pages.

Engel, et al., "High-Quality Pre-integrated Volume Rendering Using Hardware-Accelerated Pixel Shading", In Proceedings of the ACM SIGGRAPH/Eurographics Workshop on Graphics Hardware., Aug. 2001, 9 Pages.

Engel, et al., "Real-Time Volume Graphics", In Book Real-Time Volume Graphics, Jul. 21, 2006, 19 Pages.

Escalera, et al., "Automatic Chessboard Detection for Intrinsic and Extrinsic Camera Parameter Calibration", In Sensors, vol. 10, Issue 3., Mar. 2010, 18 Pages.

Fanelli, et al., "Real Time Head Pose Estimation with Random Regression Forests", In IEEE Conference on Computer Vision and Pattern Recognition, Jun. 20, 2011, 8 Pages.

Fenchel, "Automatic Labeling of Anatomical Structures in MR FastView Images Using a Statistical Atlas", In Proceedings of the 11th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), New York, NY., Sep. 6, 2008, 9 Pages.

Fisher, et al., "Virtual Environment Display System", In Proceedings of the ACM Workshop on Interactive 3D graphics, Oct. 23, 1986, 12 Pages.

Fitzgibbon, "Robust Registration of 2D and 3D Point Sets", In Proceedings of Image and Vision Computing, 2003, Dec. 1, 2003, 10 Pages.

Forahobby, "How You Become the Controller with Xbox 360 Kinect", Retrieved from: <https://web.archive.org/web/20170614002544/http://www.360-hq.com/article4096.html>, Jan. 11, 2011.

Fraham, et al., "Building Rome on a Cloudless Day", In Proceedings of European Conference on Computer Vision: Part IV, Sep. 2010, 14 Pages.

Freeman, et al., "Television Control by Hand Gestures", Retrieved from <http://www.merl.com/publications/docs/TR94-24.pdf>, Dec. 1994, 7 Pages.

Freiman, et al., "Classification of Suspected Liver Metastases Using fMRI Images: A Machine Learning Approach", In International Conference on Medical Image Computing and Computer-Assisted Intervention., Sep. 6, 2008, 94 Pages.

Fujii, et al., "Three-dimensional finger tracking using direct and reflected infrared images", In Proceedings of ACM User Interface Software and Technology, Oct. 2002, 2 Pages.

Furukawa, et al., "Towards Internet-scale Multi-view Stereo", In Proceedings of Conference on Computer Vision and Pattern Recognition, Jun. 2010, 8 Pages.

Gall, et al., "Hough Forests for Object Detection, Tracking, and Action Recognition", In IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 33, Issue 11, Nov. 2011, 15 Pages.

Garstka, et al., "View-dependent 3D Projection using Depth-Image-based Head Tracking", In Proceedings of IEEE International Workshop on Projector-Camera Systems, Jun. 24, 2011, 7 Pages.

Gee, et al., "6D Relocalisation for RGBD Cameras Using Synthetic View Regression", In Proceeding of British Machine Vision Conference, Sep. 3, 2012, 11 Pages.

"Non-Final Rejection Received in U.S. Appl. No. 13/017,518", dated Mar. 5, 2015, 19 Pages.

"Office Action Issued in Chinese Patent Application No. 201210021583.7", dated Apr. 20, 2016, 9 Pages.

"Notice of Re-Examination issued in Chinese Patent Application No. 201210021583.7", dated Jan. 12, 2017, 7 Pages.

"Fourth Office Action Issued in Chinese Patent Application No. 201210021583.7", dated Oct. 9, 2015, 3 Pages.

"Non-Final Office Action Issued in U.S. Appl. No. 13/017,729", dated Jan. 18, 2013, 18 Pages.

"Final Office Action Issued in U.S. Appl. No. 13/154,288", dated Jun. 10, 2014, 22 Pages.

"Non-Final Office Action Issued in U.S. Appl. No. 13/154,288", dated Oct. 24, 2013, 18 Pages.

"Non Final Office Action Issued in U.S. Appl. No. 13/300,542", dated Sep. 4, 2013, 10 Pages.

"Final Office Action Issued in U.S. Appl. No. 13/729,324", dated Nov. 13, 2015, 16 Pages.

"Final Office Action Issued in U.S. Appl. No. 13/729,324", dated Nov. 2, 2016, 18 Pages.

"Non Final Office Action Issued in U.S. Appl. No. 13/729,324", dated Jun. 10, 2016, 16 Pages.

"Non-Final Office Action Issued in U.S. Appl. No. 13/729,324", dated May 10, 2016, 16 Pages.

"Non-Final Office Action Issued in U.S. Appl. No. 13/729,324", dated Jun. 2, 2015, 15 Pages.

"Non-Final Office Action Issued in U.S. Appl. No. 13/729,324", dated Apr. 6, 2017, 5 Pages.

"Final Office Action Issued in U.S. Appl. No. 13/749,497", dated Nov. 10, 2014, 14 Pages.



(56)

**References Cited**

## OTHER PUBLICATIONS

“Non Final Office Action Issued in U.S. Appl. No. 13/749,497”, dated Jun. 10, 2014, 15 Pages.

“Non Final Office Action Issued in U.S. Appl. No. 13/749,497”, dated Mar. 5, 2015, 14 Pages.

“First Office Action Issued in Chinese Patent Application No. 201210021583.7”, dated Jan. 20, 2014, 14 Pages.

“Third Office Action Issued in Chinese Patent Application No. 201210021582.2”, dated Mar. 9, 2015, 7 Pages.

“Final Office Action Issued in U.S. Appl. No. 13/774,145”, dated Mar. 20, 2015, 16 Pages.

“Final Office Action Issued in U.S. Appl. No. 13/774,145”, dated Dec. 15, 2016, 18 Pages.

“Final Office Action Issued in U.S. Appl. No. 13/774,145”, dated May 3, 2016, 18 Pages.

“Non Final Office Action Issued in U.S. Appl. No. 13/774,145”, dated May 19, 2017, 20 Pages.

“Non-Final Office Action Issued in U.S. Appl. No. 13/774,145”, dated Nov. 6, 2014, 12 Pages.

“Non-Final Office Action Issued in U.S. Appl. No. 13/774,145”, dated Aug. 24, 2016, 16 Pages.

“Non-Final Office Action Issued in U.S. Appl. No. 13/774,145”, dated Oct. 2, 2015, 15 Pages.

“Non Final Office Action Issued in U.S. Appl. No. 13/902,506”, dated Mar. 13, 2015, 6 Pages.

“Non-Final Office Action Issued in U.S. Appl. No. 14/978,123”, dated Jan. 24, 2018, 10 Pages.

“Non-Final Office Action Issued in U.S. Appl. No. 14/978,123”, dated Jul. 6, 2017, 7 Pages.

“Office Action Issued in European Patent Application No. 14704433.3”, dated May 31, 2015, 4 Pages.

“Office Action Issued in European Patent Application No. 14709030.2”, dated Sep. 29, 2017, 5 Pages.

“Office Action Issued in European Patent Application No. 14709030.2”, dated Feb. 24, 2017, 6 Pages.

“Non-Final Office Action Issued in U.S. Appl. No. 15/001,645”, dated Oct. 19, 2017, 6 Pages.

“First Office Action Issued in Chinese Patent Application No. 201110118235.7”, dated Dec. 4, 2012, 8 Pages.

“First Office Action and Search Report Issued in Chinese Patent Application No. 200980149038.2”, dated Dec. 4, 2013, 10 Pages.

“First Office Action Issued in Chinese Patent Application No. 201210021228.X”, dated Dec. 13, 2013, 16 Pages.

“Fourth Office Action issued in Chinese Patent Application No. 201210021228.X”, dated Jun. 24, 2015, 7 Pages.

“Second Office action Issued in Chinese Patent Application No. 201210021228.X”, dated Jul. 8, 2014, 14 Pages.

“Third Office Action Issued in Chinese Patent Application No. 201210021228.X”, dated Dec. 12, 2014, 7 Pages.

“First Office Action and Search Report Issued in Chinese Patent Application No. 201210021241.5”, dated Dec. 24, 2013, 17 Pages.

“Second Office Action Issued in Chinese Patent Application No. 201210021241.5”, dated Aug. 7, 2014, 14 Pages.

“First Office Action Issued in Chinese Patent Application No. 201210021579.0”, dated Jul. 2, 2014, 14 Pages.

“Office Action Issued in Chinese Patent Application No. 201210021582.2”, dated Sep. 10, 2014, 7 Pages.

“Office Action Issued in Chinese Patent Application No. 201210021582.2”, dated Jan. 3, 2014, 14 Pages.

“Non Final Office Action Issued in U.S. Appl. No. 15/846,180”, dated Jan. 14, 2019, 18 Pages.

Westermann, “Efficiently Using Graphics Hardware in Volume Rendering Applications”, In the Proceedings of the 25th Annual Conference on Computer Graphics and Interactive Techniques, ACM New York, NY, USA., Jul. 24, 1998, 9 Pages.

Whelan, et al., “Robust Tracking for Real-Time Dense RGB-D Mapping with Kintinuous”, In technical report of MIT., Sep. 2012, 10 Pages.

Williams, et al., “Automatic Relocalization and Loop Closing for Real-Time Monocular SLAM”, In IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 33, Issue 9, Sep. 2011, 14 Pages.

Williams, et al., “Real-Time SLAM Relocalisation”, In Proceedings of International Conference on Computer Vision (ICCV), Rio de Janeiro, Brazil, Oct. 2007, 8 Pages.

Wilson, et al., “Combining Multiple Depth Cameras and Projectors for Interactions On, Above, and Between Surfaces”, In Proceedings of the 23rd Annual ACM Symposium on User Interface Software and Technology, Oct. 3, 2010, 10 Pages.

Winder, et al., “Learning Local Image Descriptors”, In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Jul. 17, 2007, 8 Pages.

Wren, et al., “Pfinder: Real-Time Tracking of the Human Body”, In IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 19, Issue 7, Jul. 1997, 6 Pages.

Wu, et al., “3D Model Matching with Viewpoint-Invariant Patches (VIP)”, In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Jun. 23, 2008, 8 Pages.

Wuhrer, et al., “Human Shape Correspondence with Automatically Predicted Landmarks”, In Journal of Machine Vision and Applications, vol. 22, Aug. 6, 2011, 9 Pages.

Wurm, et al., “OctoMap: A Probabilistic, Flexible, and Compact 3D Map Representation for Robotic Systems”, In Proceedings of Workshop on Best Practice in 3D Perception and Modeling for Mobile Manipulation (ICRA), May 2010, 8 Pages.

Xia, et al., “Human Detection Using Depth Information by Kinect”, In Proceeding of IEEE Computer Vision and Pattern Recognition, Published on: Jun. 21, 2011, 8 Pages.

Yagel, “Volume Viewing Algorithms: Survey”, In International Spring School on Visualization., 2000, 22 Pages.

Yan, et al., “Robustly Aligning a Shape Model and Its Application to Car Alignment of Unknown Pose”, In IEEE Transactions on Pattern Analysis and Machine Intelligence, IEEE Computer Society, USA, vol. 33, No. 9., Sep. 1, 2011, pp. 1860-1876.

Yao, et al., “Simultaneous Location Detection of Multi-Organ by Atlas-Guided Eigen-organ method in Volumetric Medical Images”, In International Journal of Computer Assisted Radiology and Surgery, vol. 1., 2006, 9 pages.

Yi, et al., “Discriminative, Semantic Segmentation of Brain Tissue in MR Images”, In International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, Berlin, Heidelberg., Sep. 20, 2009, 8 Pages.

Yin, et al., “Tree-based Classifiers for Bilayer Video Segmentation”, In the Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Jun. 17, 2007, 8 Pages.

Yu, et al., “Monocular Video Foreground/Background Segmentation by Tracking Spatial-Color Gaussian Mixture Models”, In Proceedings of Workshop on Motion and Video Computing, Feb. 2007, 8 Pages.

Zach, et al., “A Globally Optimal Algorithm for Robust TV-L1 Range Image Integration”, In IEEE 11th International Conference on Computer Vision, Oct. 2007, 2007, 8 Pages.

Zambal, et al., “Entropy-Optimized Texture Models”, In Proceedings of the 11th International Conference on Medical Image Computing and Computer-Assisted Intervention, New York, Sep. 6, 2008, 9 Pages.

Zhan, et al., “Active Scheduling of Organ Detection and Segmentation in Whole-Body Medical Images”, In Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, Berlin, Heidelberg., Sep. 6, 2008, pp. 313-321.

Zhao, Liang, “Dressed Human Modeling, Detection, and Parts Localization”, A Dissertation Submitted in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy at The Robotics Institute Carnegie Mellon University Pittsburgh., Jul. 26, 2001, 121 Pages.

Zhong, Yuan, “A Weighting Scheme for Content-Based Image Retrieval”, Retrieved From <http://www.library.uow.edu.au/adt-NWU/uploads/approved/adt-NWU20080116.091123/public/01Front.pdf>, 11 Pages.



(56)

**References Cited**

## OTHER PUBLICATIONS

Zhou, et al., "Data-Parallel Octrees for Surface Reconstruction", In Proceedings of IEEE Transactions on Visualization and Computer Graphics, vol. 17, Issue 5, May 27, 2011, 13 Pages.

Zhou, et al., "Highly Parallel Surface Reconstruction", In Microsoft Research, Technical Report MSR-TR-2008-53., Apr. 2008, 10 Pages.

Zhuang, et al., "An Atlas-based Segmentation Propagation Framework Using Locally Affine Registration", In Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention, Sep. 6, 2008, pp. 425-433.

Ziegler, et al., "Tracking of the Articulated Upper Body On Multi-View Stereo Image Sequences", In IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 1., Jun. 17, 2006, pp. 774-781.

Zillner, et al., "Semantics and CBIR: A Medical Imaging Perspective", In Proceedings of the 2008 international conference on Content-based image and video retrieval, 2008., Jul. 7, 2008, 10 Pages.

Newcombe, et al., "Three Dimensional Environment Reconstruction", Retrieved from: <https://patents.google.com/patent/US8587583B2/en>, Nov. 19, 2013, 34 Pages.

Nguyen., "Gpu Gems 3. Addison-Wesley Professional", Retrieved From: [http://http.developer.nvidia.com/GPUGems3/gpugems3\\_pref01.html](http://http.developer.nvidia.com/GPUGems3/gpugems3_pref01.html), 2007, 4 Pages.

Ni, et al., "Epitomic Location Recognition", In Proceeding of IEEE Conference on Computer Vision and Pattern Recognition, Jun. 23, 2008, 8 Pages.

Nister, David, "Preemptive RANSAC for Live Structure and Motion Estimation", In Proceedings of the Ninth IEEE International Conference on Computer Vision, Oct. 13, 2003, 8 Pages.

Nister, et al., "Scalable Recognition with a Vocabulary Tree", In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Jun. 17, 2006, 8 Pages.

Osher, et al., "Level Set Methods and Dynamic Implicit Surfaces, Signed Distance Functions", In Book Level Set Methods and Dynamic Implicit Surfaces, Series Title Applied Mathematical Sciences, Series vol. 153, Apr. 6, 2006, 7 Pages.

P120100292, "Office Action Issued in Argentina Patent Application No. P120100292", dated Dec. 26, 2016, 4 Pages.

Parker, et al., "Interactive Ray Tracing for Isosurface Rendering", In Proceedings of IEEE Computer Society, Conference on Visualization, Oct. 1998, 6 Pages.

Pascal, et al., "Vision Based 3D Tracking and Pose Estimation for Mixed Reality", In Emerging Technologies of Augmented Reality: Interfaces and Design, 2007, 20 Pages.

Paul, et al., "A Method for Registration of 3-D Shapes", In Proceedings of IEEE Transactions on Pattern Analysis and Machine Intelligence—Special issue on interpretation of 3-D scenes—part-II, vol. 14, Issue 2, Feb. 1, 1992, 18 Pages.

Pavlovic, et al., "Visual Interpretation of Hand Gestures for Human-Computer Interaction: A Review", In Proceedings of the IEEE Transactions on Pattern Analysis and Machine Intelligence vol. 19, No. 7, Jul. 1, 1997, 19 Pages.

"International Search Report and Written Opinion Issued in PCT Application No. PCT/US12/20681", dated Aug. 28, 2012, 8 Pages.

"International Search Report and Written Opinion Issued in PCT Application No. PCT/US12/20686", dated Aug. 22, 2012, 9 Pages.

"International Search Report and Written Opinion Issued in PCT Application No. PCT/US12/20687", dated Aug. 30, 2012, 8 Pages.

"International Search Report and Written Opinion Issued in PCT Application No. PCT/US12/62875", dated Mar. 25, 2013, 16 Pages.

"International Search Report and Written Opinion Issued in PCT Application No. PCT/US2012/020687", dated Aug. 30, 2012, 8 Pages.

"International Search Report & Written Opinion Received For PCT Application No. PCT/US2013/077736", dated Mar. 27, 2014, 15 Pages.

"International Preliminary Report on Patentability Issued in PCT Application No. PCT/US2014/012226", dated Jul. 9, 2015, 6 Pages.

"International Search Report & Written Opinion Issued in PCT Application No. PCT/US2014/012226", dated May 12, 2014, 8 Pages.

"Second Written Opinion Issued in PCT Application No. PCT/US2014/012226", dated Apr. 7, 2015, 5 Pages.

"International Preliminary Report on Patentability Issued in PCT Application No. PCT/US2014/016749", dated Jun. 5, 2015, 13 Pages.

"International Search Report & Written Opinion Issued in PCT Patent Application No. PCT/US2014/016749", dated May 12, 2014, 8 Pages.

"Second Written Opinion Issued in PCT Application No. PCT/US2014/016749", dated Aug. 21, 2014, 6 Pages.

Persson, et al., "Volume Rendering of Three-Dimensional Drip Infusion CT Cholangiography in Patients with Suspected Obstructive Biliary Disease: a Retrospective Study", In The British Journal of Radiology, 78.936., Dec. 2005, 21 Pages.

Pescia, et al., "Automatic Detection of Liver Tumors", In 5th IEEE International Symposium on Biomedical Imaging, May 14, 2008, 5 Pages.

Pohl, et al., "A Hierarchical Algorithm for MR Brain Image Parcellation", Retrieved from :<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2768067/>, Sep. 26, 2007, 22 Pages.

Pollefeys, "Detailed Real-Time Urban 3D Reconstruction from Video", In International Journal of Computer Vision, vol. 78 Issue 2-3., Jul. 2008, 43 Pages.

Poppe, et al., "Body-part templates for recovery of 2D human poses under occlusion", In the Proceedings of International Conference on Articulated Motion and Deformable Objects, Springer, Berlin, Heidelberg., Jul. 9, 2008, 10 Pages.

Prasad, et al., "Multi-level Classification of Emphysema in HRCT Lung Images Using Delegated Classifiers", In Proceedings of the 11th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), New York, NY, Sep. 6, 2008, 9 Pages.

Priem, et al., "Visualization in Medicine", In Book Visualization in Medicine: Theory, Algorithms, and Applications, Morgan Kaufmann Publishers Inc. San Francisco, CA, USA, Jun. 21, 2007, 680 Pages.

Purcell, et al., "Ray Tracing on Programmable Graphics Hardware", Proceedings of the 29th Annual Conference on Computer Graphics and Interactive Techniques, vol. 1, Issue 3, Jun. 23, 2002, 10 Pages.

Qian, et al., "A Gesture-Driven Multimodal Interactive Dance System", In Proceedings of the IEEE International Conference on Multimedia and Expo, vol. 3, Jun. 27, 2004, pp. 1579-1582.

Quinlan, "C4.5: Programs for Machine Learning", In Book Machine Learning, Morgan Kaufmann Publishers, San Francisco, CA, 1993.

Raca, Mirko, "Improving 3D Human Pose Estimation", Retrieved from : [https://wiki.epfl.ch/edicpublic/documents/Candidacy%20exam/candidacy\\_edic\\_raca.pdf](https://wiki.epfl.ch/edicpublic/documents/Candidacy%20exam/candidacy_edic_raca.pdf), Jul. 4, 2011, 8 Pages.

Rautek, "Semantic Layers for Illustrative Volume Rendering", In the proceedings of IEEE Transactions on Visualization and Computer Graphics, vol. 13(6)., Nov. 2007, 8 Pages.

Robler, et al., "Dynamic Shader Generation for Flexible Multi-Volume Visualization", IEEE, Pacific Visualization Symposium (PacificVis), Mar. 5, 2008, 8 Pages.

Robler, et al., "GPU-based Multi-Volume Rendering for the Visualization of Functional Brain Images", In SimVis, Jan. 2006, 14 Pages.

Rosenhahn, et al., "Automatic Human Model Generation", In International Conference on Computer Analysis of Images and Patterns, Sep. 5, 2005, pp. 41-48.

Rosten, et al., "Faster and Better: A Machine Learning Approach to Corner Detection", In IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 32, Issue 1, Jan. 2010, 15 Pages.

Rubin, et al., "Medical Imaging on the Semantic Web: Annotation and Image Markup.", In AAAI Spring Symposium: Semantic Scientific Knowledge Integration, Mar. 26, 2008, 6 Pages.

Ruble, et al., "ORB: An Efficient Alternative to SIFT or SURF", In Proceeding of IEEE International Conference on Computer Vision, Nov. 6, 2011, 8 Pages.

Rusinkiewicz, et al., "Efficient variants of the ICP algorithm", In Third International Conference on 3-D Digital Imaging and Modeling, 2001, pp. 145-152.



(56)

**References Cited**

## OTHER PUBLICATIONS

Rusinkiewicz, et al., "Real-Time 3D Model Acquisition", In Proceedings of the 29th Annual Conference on Computer Graphics and Interactive Techniques, Jul. 2002, 9 Pages.

Salama, et al., "Interactive vol. Rendering on Standard PC Graphics Hardware Using Multi-Textures and Multi-Stage Rasterization", In the Proceedings of the ACM SIGGRAPH/Eurographics Workshop on Graphics Hardware, Interlaken, Switzerland., Aug. 21-22, 2000, pp. 109-118.

Sato, et al., "Tissue Classification Based on 3D Local Intensity Structures for vol. Rendering", In Journal IEEE Transactions on Visualization and Computer Graphics archive vol. 6 Issue 2., Apr. 6, 2000, p. 160-180.

Sattler, et al., "Fast Image-Based Localization Using Direct 2D-to-3D Matching", In Proceeding of International Conference on Computer Vision, Nov. 6, 2011, 8 Pages.

Schindler, et al., "City-Scale Location Recognition", In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Jun. 17, 2007, 7 Pages.

Schmalstieg, "Crepuscular Rays for Tumor Accessibility Planning", In Journal IEEE Transactions on Visualization and Computer Graphics Archive, vol. 17, Issue 12., Dec. 2011, pp. 2163-2172.

Se, et al., "Vision-Based Global Localization and Mapping for Mobile Robots", In Journal of IEEE Transaction on Robotics, vol. 21, Issue 3, Jun. 2005, 12 Pages.

Seitz, Steven M., et al., "A Comparison and Evaluation of Multiview Stereo Reconstruction Algorithms", Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06), Jun. 17, 2006, 8 Pages.

"Final Office Action Issued in U.S. Appl. No. 15/846,180", dated Jul. 15, 2019, 23 Pages.

"Office Action Issued in European Patent Application No. 12742358.0", dated Oct. 2, 2019, 6 Pages.

Fransen et al. "Real-Time Face and Object Tracking", In Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, Oct. 11, 2009, pp. 2483-2488.

"Notice on Reexamination Issued in Chinese Patent Application No. 201380068406.7", dated Sep. 2, 2019, 7 Pages.

"Non Final Office Action Issued In U.S. Appl. No. 15/846,180", dated Dec. 23, 2019, 26 Pages.

"Final Office Action Issued in U.S. Appl. No. 15/846,180", dated May 20, 2020, 29 Pages.

"Decision on Reexamination Issued in Chinese Patent Application No. 201380068406.7", dated Jan. 23, 2020, 11 Pages.

"Non-Final Office Action Issued in U.S. Appl. No. 15/846,180", dated Oct. 7, 2020, 34 Pages.

"Office Action Issued in European Patent Application No. 12191941.9", dated Nov. 19, 2020, 7 Pages.

"Final Office Action Issued in U.S. Appl. No. 15/846,180", dated Apr. 13, 2021, 33 Pages.

"Office Action Issued in Argentina Patent Application No. P120100266", dated Oct. 21, 2021, 9 Pages.

"Notice of Allowance Issued in U.S. Appl. No. 15/846,180", dated Aug. 31, 2021, 9 Pages.

"Office Action Issued in European Patent Application No. 12191941.9", dated Jun. 28, 2022, 7 Pages.

"Non Final Office Action Issued in U.S. Appl. No. 17/565,344", dated Sep. 12, 2022, 8 Pages.

"Office Action Issued in Argentina Patent Application No. P120100266", dated Aug. 24, 2022, 4 Pages.

\* cited by examiner



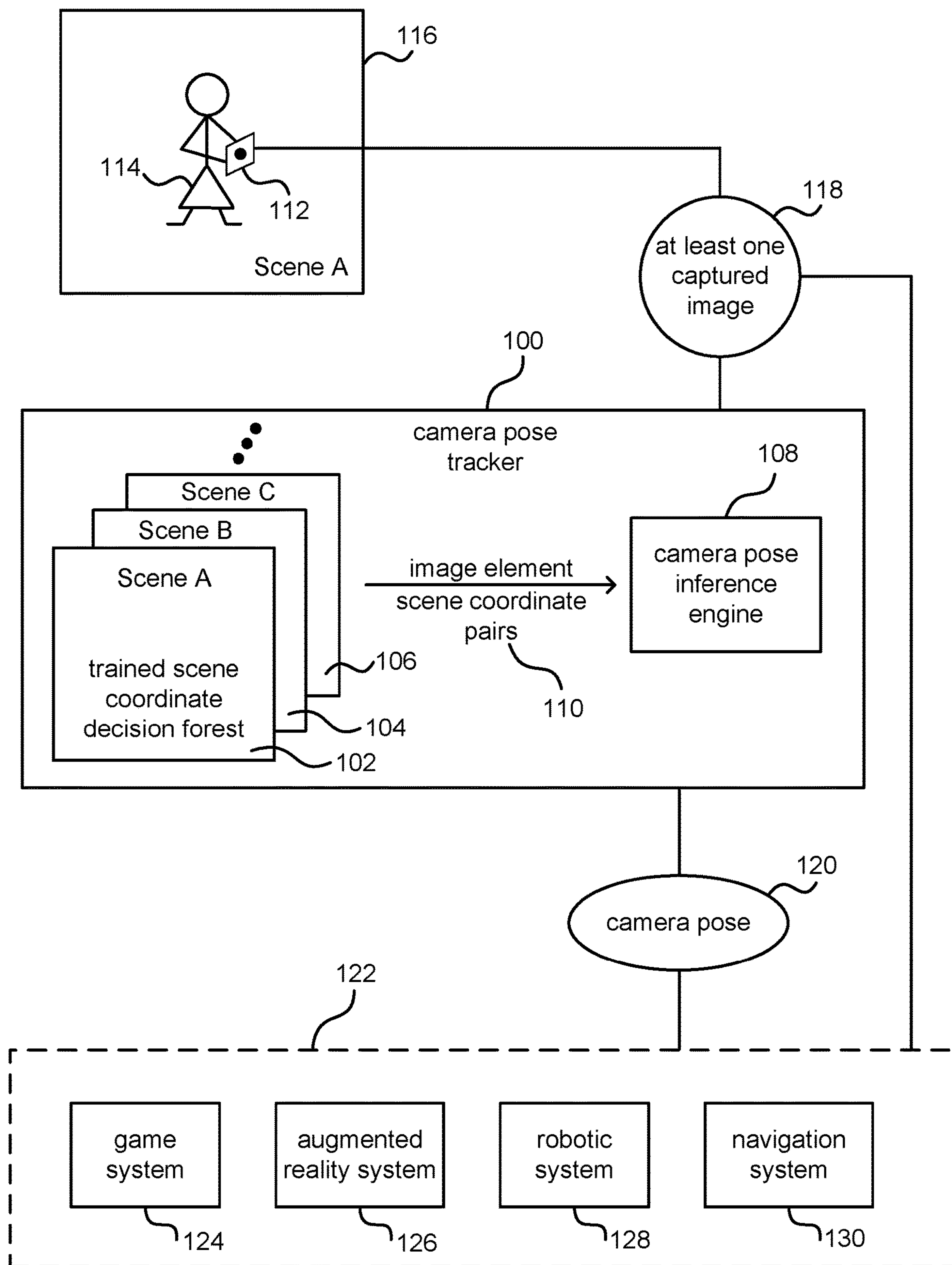


FIG. 1



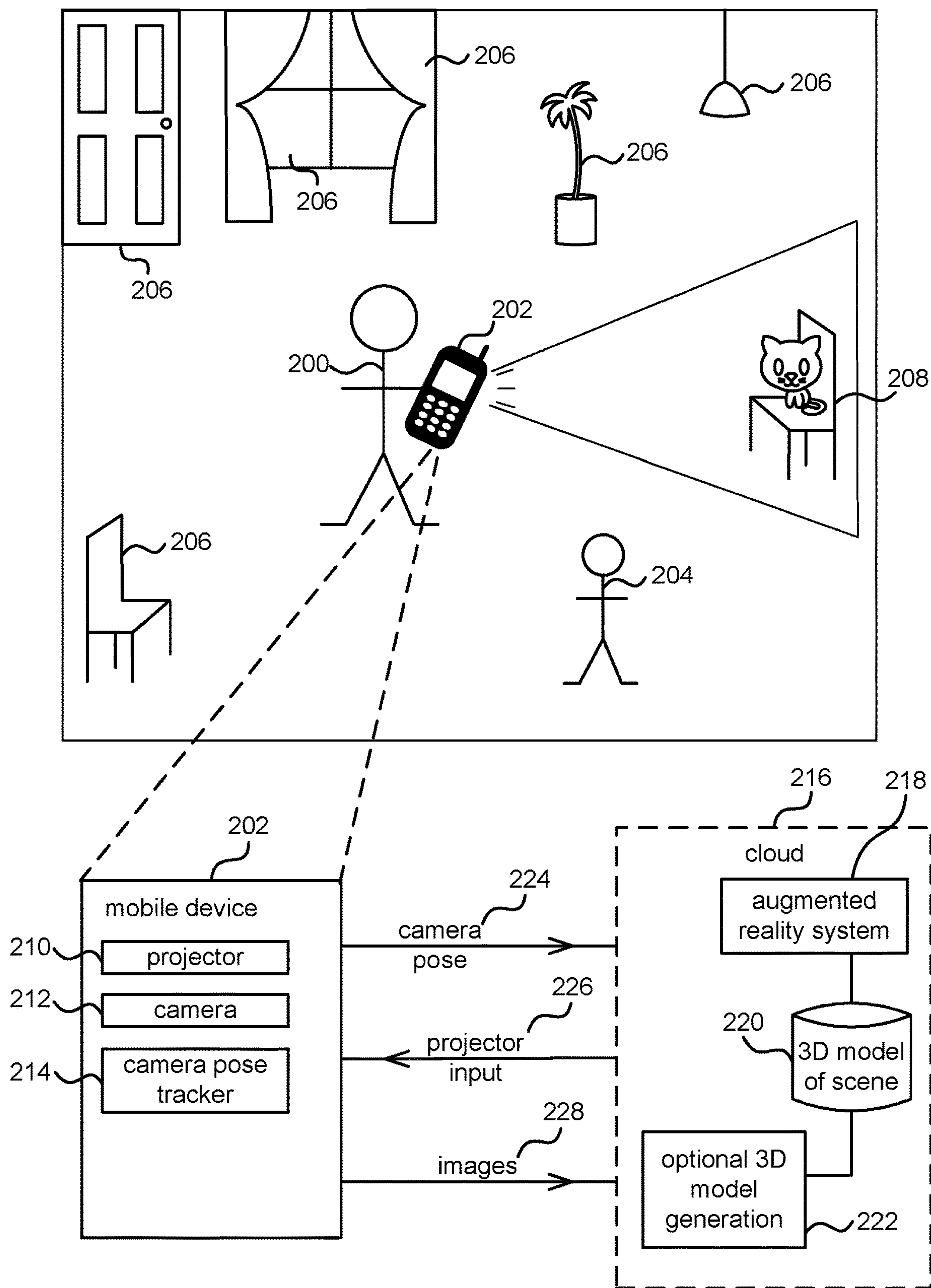


FIG. 2



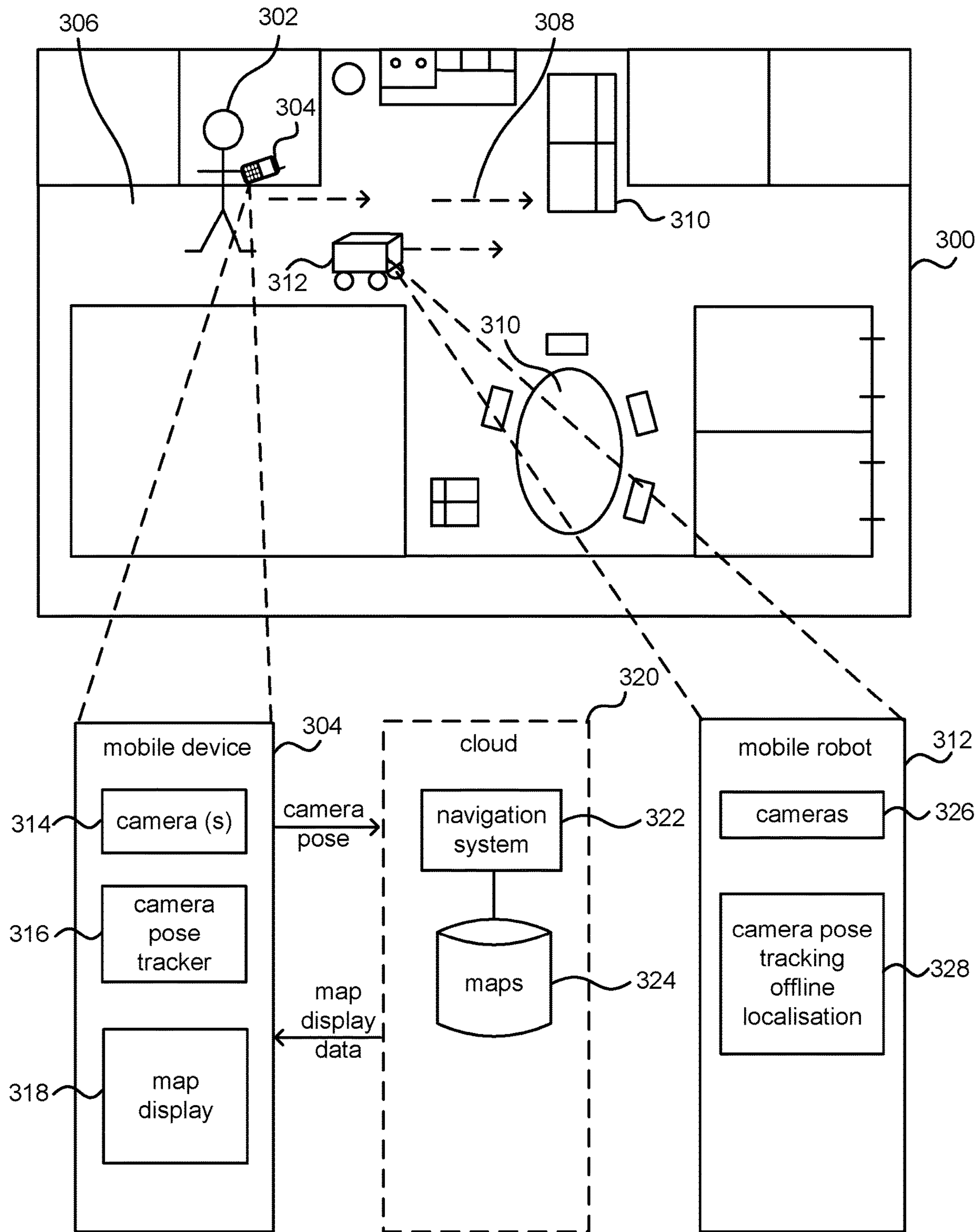


FIG. 3



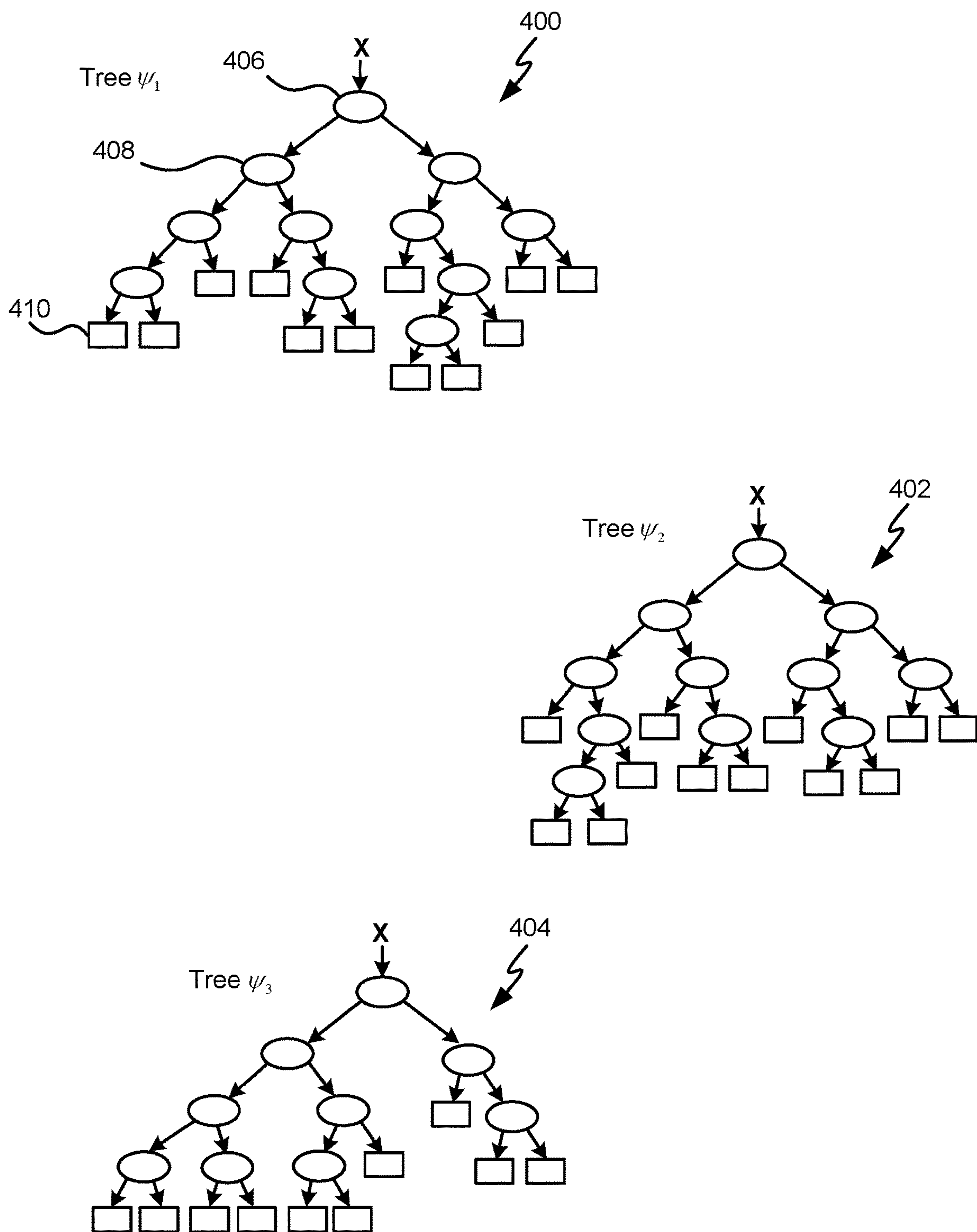


FIG. 4



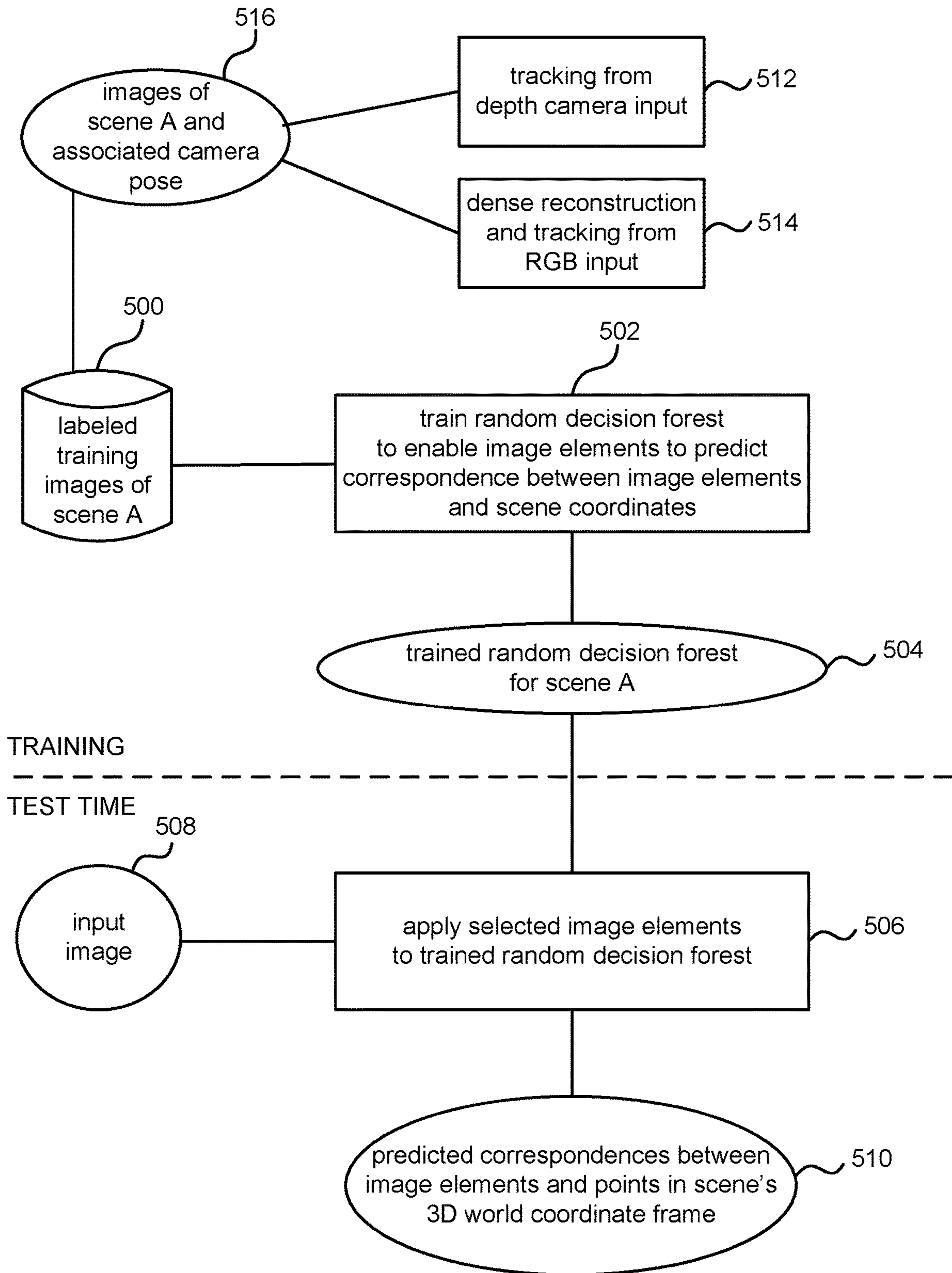


FIG. 5



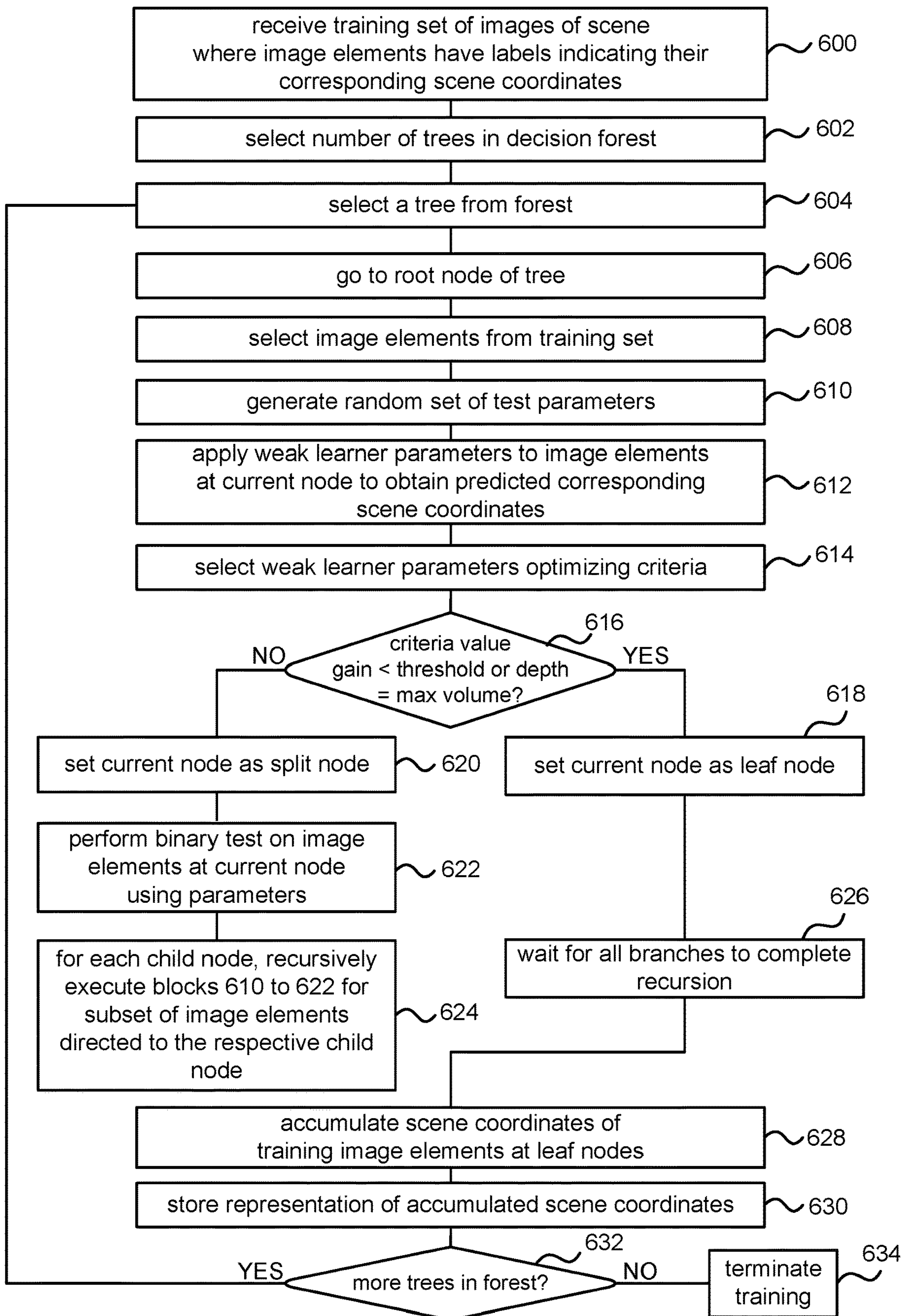


FIG. 6



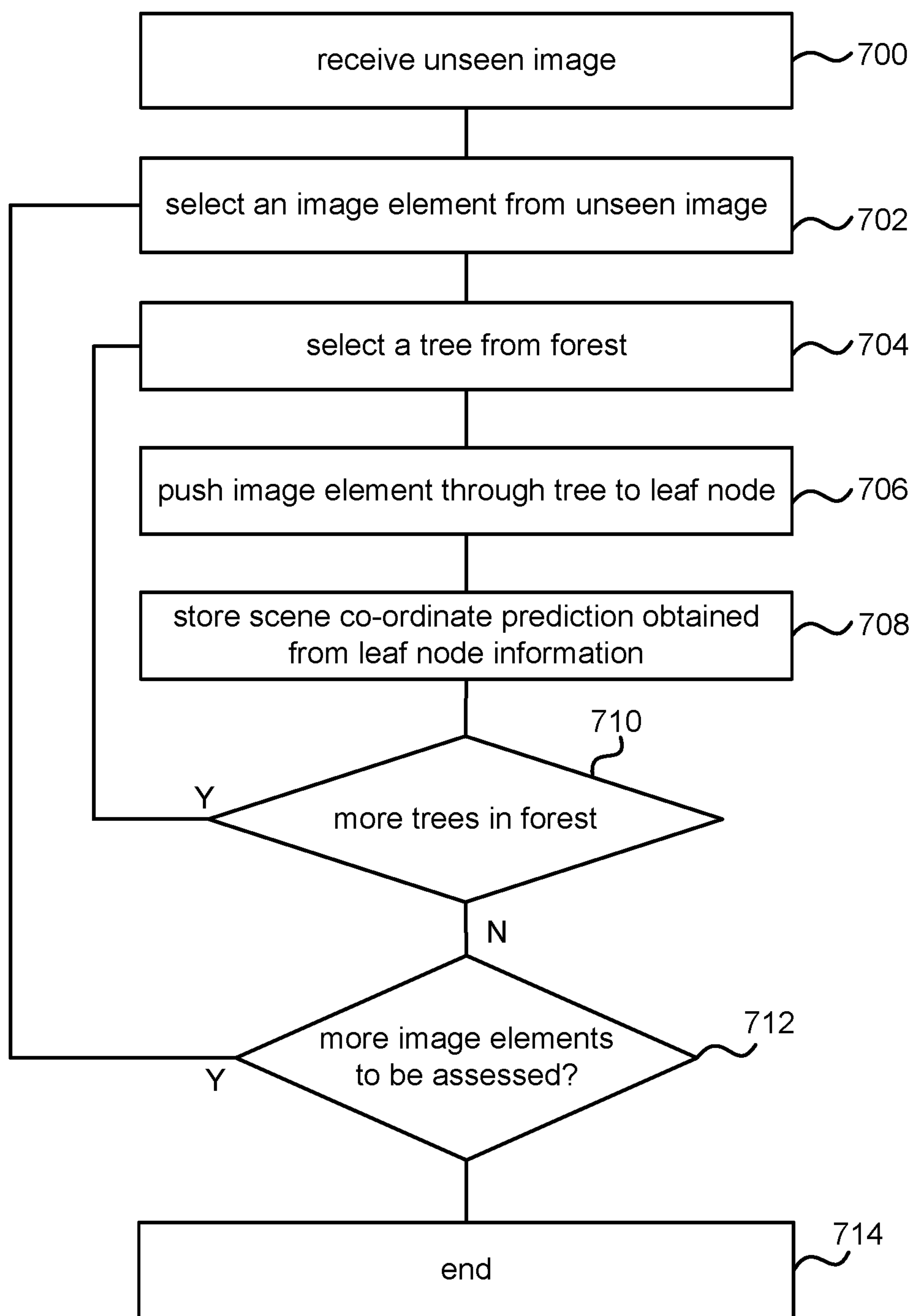


FIG. 7



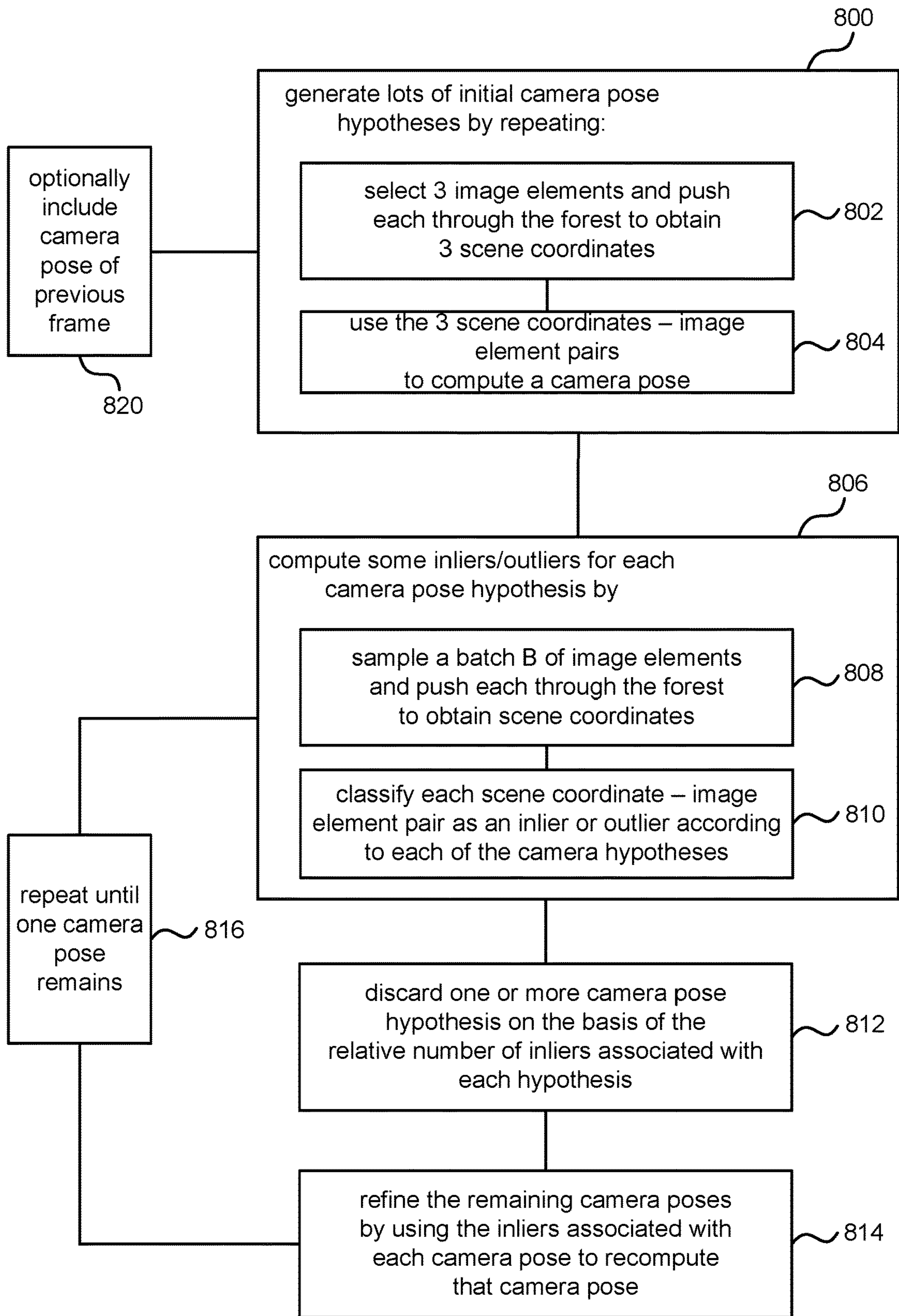


FIG. 8



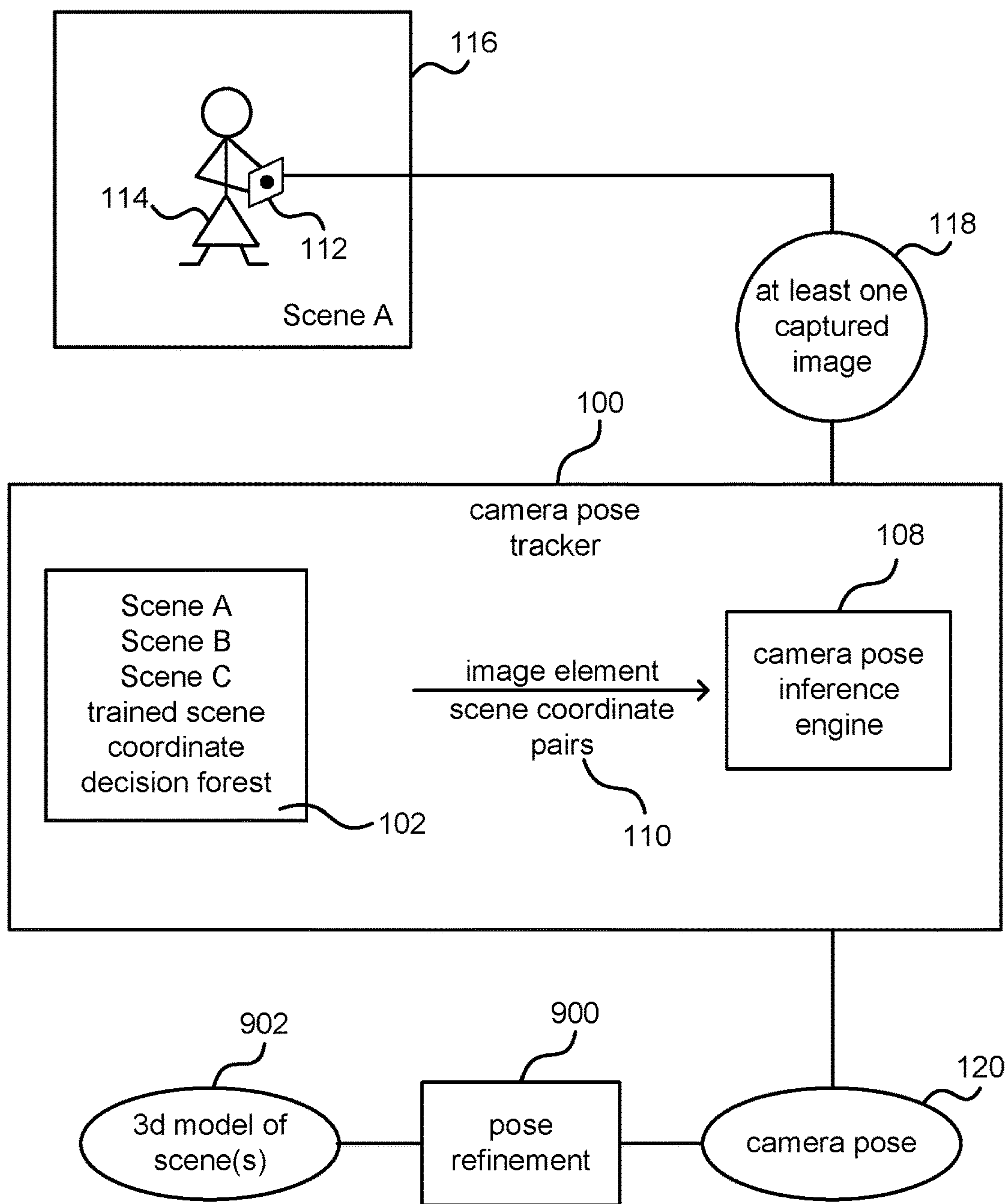


FIG. 9



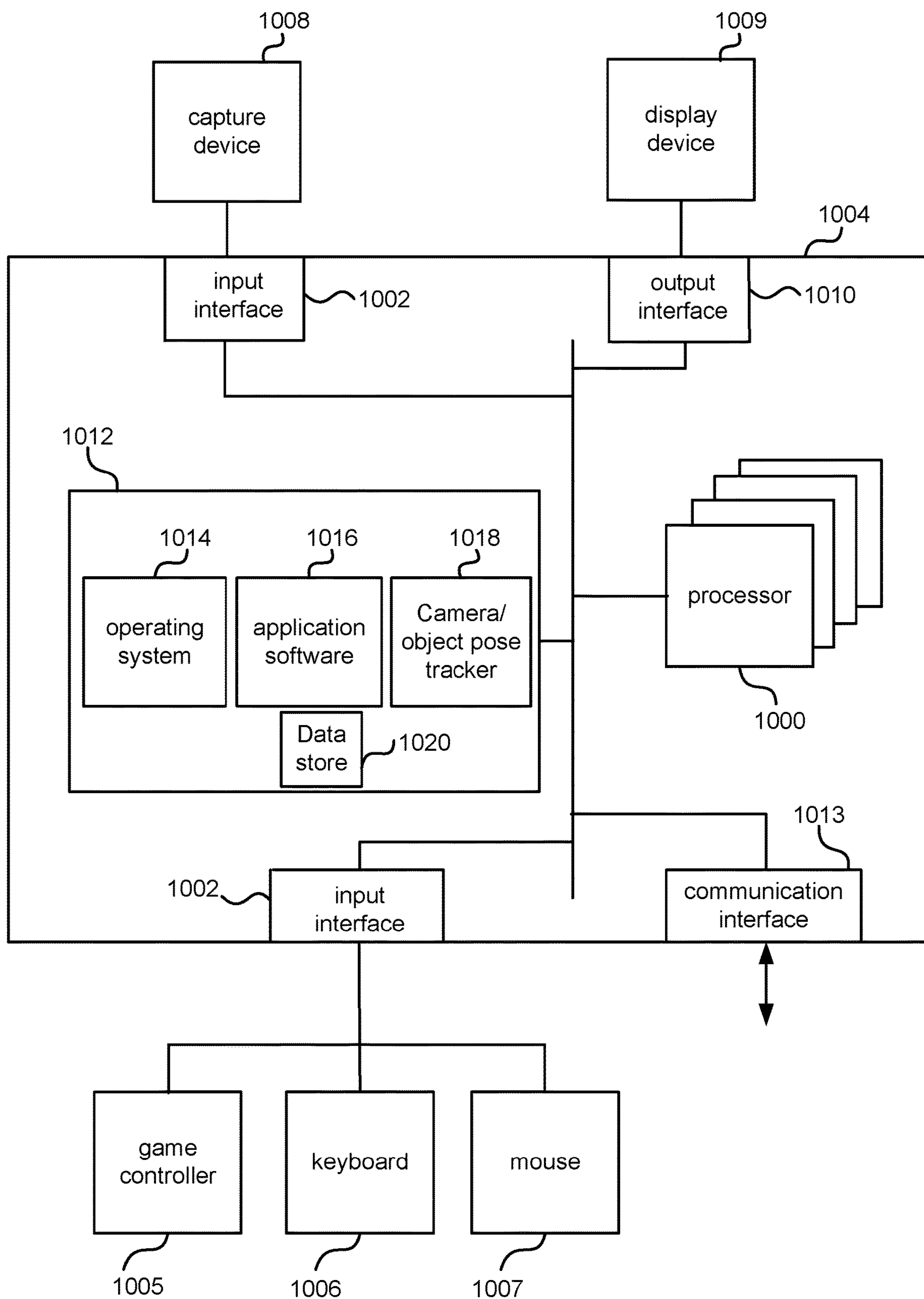


FIG. 10



**1****CAMERA/OBJECT POSE FROM PREDICTED  
COORDINATES****CROSS-REFERENCE TO RELATED  
APPLICATION**

This application is a continuation of U.S. Non-Provisional application Ser. No. 13/774,145, filed on Feb. 22, 2013, and entitled "CAMERA/OBJECT POSE FROM PREDICTED COORDINATES" and is hereby incorporated by reference herein for all intents and purposes.

**BACKGROUND**

For many applications, such as robotics, vehicle navigation, computer game applications, medical applications and other problem domains, it is valuable to be able to find orientation and position of a camera as it moves in a known environment. Orientation and position of a camera is known as camera pose and may comprise six degrees of freedom (three of translation and three of rotation). Where a camera is fixed and an object moves relative to the camera it is also useful to be able to compute the pose of the object.

A previous approach uses keyframe matching where a whole test image is matched against exemplar training images (keyframes). *K* matching keyframes are found, and the poses (keyposes) of those keyframes are interpolated to generate an output camera pose. Keyframe matching tends to be very approximate in the pose result.

Another previous approach uses keypoint matching where a sparse set of interest points are detected in a test image and matched using keypoint descriptors to a known database of descriptors. Given a putative set of matches, a robust optimization is run to find the camera pose for which the largest number of those matches are consistent geometrically. Keypoint matching struggles in situations where too few keypoints are detected.

Existing approaches are limited in accuracy, robustness and speed.

The embodiments described below are not limited to implementations which solve any or all of the disadvantages of known systems for finding camera or object pose.

**SUMMARY**

The following presents a simplified summary of the disclosure in order to provide a basic understanding to the reader. This summary is not an extensive overview of the disclosure and it does not identify key/critical elements or delineate the scope of the specification. Its sole purpose is to present a selection of concepts disclosed herein in a simplified form as a prelude to the more detailed description that is presented later.

Camera or object pose calculation is described, for example, to relocalize a mobile camera (such as on a smart phone) in a known environment or to compute the pose of an object moving relative to a fixed camera. The pose information is useful for robotics, augmented reality, navigation and other applications. In various embodiments where camera pose is calculated, a trained machine learning system associates image elements from an image of a scene, with points in the scene's 3D world coordinate frame. In examples where the camera is fixed and the pose of an object is to be calculated, the trained machine learning system associates image elements from an image of the object with points in an object coordinate frame. In examples, the image

**2**

elements may be noisy and incomplete and a pose inference engine calculates an accurate estimate of the pose.

Many of the attendant features will be more readily appreciated as the same becomes better understood by reference to the following detailed description considered in connection with the accompanying drawings.

**DESCRIPTION OF THE DRAWINGS**

The present description will be better understood from the following detailed description read in light of the accompanying drawings, wherein:

FIG. 1 is a schematic diagram of a camera pose tracker for relocalizing a mobile camera (such as in a smart phone) in scene A;

FIG. 2 is a schematic diagram of a person holding a mobile device with a camera and a camera pose tracker and which communicates with an augmented reality system to enable an image of a cat to be projected into the scene in a realistic manner;

FIG. 3 is a schematic diagram of a person and a robot each with a camera and a camera pose tracker;

FIG. 4 is a schematic diagram of three random decision trees forming at least part of a random decision forest;

FIG. 5 is a flow diagram of a method of training a random decision forest to predict correspondences between image elements and scene coordinates; and using the trained random decision forest;

FIG. 6 is a flow diagram of a method of training a random decision forest using images of a scene where image elements have labels indicating their corresponding scene coordinates;

FIG. 7 is a flow diagram of a method of using a trained random decision forest to obtain scene coordinate—image element pairs;

FIG. 8 is a flow diagram of a method at a camera pose inference engine of using scene-coordinate-image element pairs to infer camera pose;

FIG. 9 is a schematic diagram of the camera pose tracker of FIG. 1 where a 3D model of the scene is available;

FIG. 10 illustrates an exemplary computing-based device in which embodiments of a camera or object pose tracker may be implemented.

Like reference numerals are used to designate like parts in the accompanying drawings.

**DETAILED DESCRIPTION**

The detailed description provided below in connection with the appended drawings is intended as a description of the present examples and is not intended to represent the only forms in which the present example may be constructed or utilized. The description sets forth the functions of the example and the sequence of steps for constructing and operating the example. However, the same or equivalent functions and sequences may be accomplished by different examples.

Although the present examples are described and illustrated herein as being implemented using a random decision forest, the system described is provided as an example and not a limitation. As those skilled in the art will appreciate, the present examples may be implemented using a variety of different types of machine learning systems including but not limited to support vector machines, Gaussian process regression systems.

FIG. 1 is a schematic diagram of a camera pose tracker for relocalizing a mobile camera (such as in a smart phone) in



scene A. In this example a person **114** is holding the mobile camera **112** which is integral with a communications device such as a smart phone. The person **114** uses the mobile camera **112** to capture at least one image **118** of scene A **116**, such as a living room, office or other environment. The image may be a depth image, a color image (referred to as an RGB image) or may comprise both a depth image and a color image. In some examples a stream of images is captured by the mobile camera.

A camera pose tracker **100** is either integral with the smart phone or is provided at another entity in communication with the smart phone. The camera pose tracker **100** is implemented using software and/or hardware as described in more detail below with reference to FIG. **10**. The camera pose tracker **100** comprises a plurality of trained scene coordinate decision forests **102**, **104**, **106** one for each of a plurality of scenes. The trained scene coordinate decision forests may be stored at the camera pose tracker or may be located at another entity which is in communication with the camera pose tracker. Each scene coordinate decision forest is a type of machine learning system which takes image elements (from images of its associated scene) as input and produces estimates of scene coordinates (in world space) of points in a scene which the image elements depict. Image elements may be pixels, groups of pixels, voxels, groups of voxels, blobs, patches or other components of an image. Other types of machine learning system may be used in place of the scene coordinate decision forest. For example, support vector machine regression systems, Gaussian process regression systems.

A decision forest comprises one or more decision trees each having a root node, a plurality of split nodes and a plurality of leaf nodes. Image elements of an image may be pushed through trees of a decision forest from the root to a leaf node in a process whereby a decision is made at each split node. The decision is made according to characteristics of the image element and characteristics of test image elements displaced therefrom by spatial offsets specified by the parameters at the split node. At a split node the image element proceeds to the next level of the tree down a branch chosen according to the results of the decision. The random decision forest may use regression or classification as described in more detail below. During training, parameter values (also referred to as features) are learnt for use at the split nodes and data is accumulated at the leaf nodes. For example, distributions of scene coordinates are accumulated at the leaf nodes.

Storing all the scene coordinates at the leaf nodes during training may be very memory intensive since large amounts of training data are typically used for practical applications. The scene coordinates may be aggregated in order that they may be stored in a compact manner. Various different aggregation processes may be used. An example in which modes of the distribution of scene coordinates are store is described in more detail below.

In the example of FIG. **1** there is a plurality of trained scene coordinate decision forests; one for each of a plurality of scenes. However, it is also possible to have a single trained scene coordinate decision forest which operates for a plurality of scenes. This is explained below with reference to FIG. **9**.

The scene coordinate decision forest(s) provide image element-scene coordinate pair estimates **110** for input to a camera pose inference engine **108** in the camera pose tracker **100**. Information about the certainty of the image element-scene coordinate estimates may also be available. The camera pose inference engine **108** may use an energy

optimization approach to find a camera pose which is a good fit to a plurality of image element—scene coordinate pairs predicted by the scene coordinate decision forest. This is described in more detail below with reference to FIG. **8**. In some examples scene coordinates for each available image element may be computed and used in the energy optimization. However, to achieve performance improvements whilst retaining accuracy, a subsample of image elements may be used to compute predicted scene coordinates.

The camera pose inference engine **108** uses many image element-scene coordinate pairs **110** to infer the pose of the mobile camera **112** using an energy optimization approach as mentioned above. Many more than three pairs (the minimum needed) may be used to improve accuracy. For example, the at least one captured image **118** may be noisy and may have missing image elements, especially where the captured image **118** is a depth image. On the other hand, to obtain a scene coordinate prediction for each image element in an image is computationally expensive and time consuming because each image element needs to be pushed through the forest as described with reference to FIG. **7**. Therefore, in some examples, the camera pose inference engine may use an iterative process which gives the benefit that a subsample of image elements are used to compute scene coordinate predictions whilst taking accuracy into account.

The camera pose **120** output by the camera pose tracker may be in the form of a set of parameters with six degrees of freedom, three indicating the rotation of the camera and three indicating the position of the camera. For example, the output of the camera pose tracker is a set of registration parameters of a transform from camera space to world space. In some examples these registration parameters are provided as a six degree of freedom (6DOF) pose estimate in the form of an  $SE_3$  matrix describing the rotation and translation of the camera relative to real-world coordinates.

The camera pose **120** output by the camera pose tracker **100** may be input to a downstream system **122** together with the captured image(s) **118**. The downstream system may be a game system **124**, an augmented reality system **126**, a robotic system **128**, a navigation system **130** or other system. An example where the downstream system **122** is an augmented reality system is described with reference to FIG. **2**.

The examples described show how camera pose may be calculated. These examples may be modified in a straightforward manner to enable pose of an object to be calculated where the camera is fixed. In this case the machine learning system is trained using training images of an object where image elements are labeled with object coordinates. An object pose tracker is then provided which uses the methods described herein adapted to the situation where the camera is fixed and pose of an object is to be calculated.

Alternatively, or in addition, the camera pose tracker or object pose tracker described herein can be performed, at least in part, by one or more hardware logic components. For example, and without limitation, illustrative types of hardware logic components that can be used include Field-programmable Gate Arrays (FPGAs), Program-specific Integrated Circuits (ASICs), Program-specific Standard Products (ASSPs), System-on-a-chip systems (SOCs), Complex Programmable Logic Devices (CPLDs), Graphics Processing Units (GPUs).

FIG. **2** is a schematic diagram of a person **200** holding a mobile device **202** which has a camera **212**, a camera pose tracker **214** and a projector **210**. For example, the mobile device may be a smart phone. Other components of the mobile device to enable it to function as a smart phone such



## 5

as a communications interface, display screen, power source and other components are not shown for clarity. A person 200 holding the mobile device 202 is able to capture images of the scene or environment in which the user is moving. In the example of FIG. 2 the scene or environment is a living room containing various objects 206 and another person 204.

The mobile device is able to communicate with one or more entities provided in the cloud 216 such as an augmented reality system 218, a 3D model of the scene 220 and an optional 3D model generation system 222.

For example, the user 200 operates the mobile device 202 to capture images of the scene which are used by the camera pose tracker 214 to compute the pose (position and orientation) of the camera. At the consent of the user, the camera pose is sent 224 to the entities in the cloud 216 optionally with the images 228. The augmented reality system 218 may have access to a 3D model of the scene 220 (for example, a 3D model of the living room) and may use the 3D model and the camera pose to calculate projector input 226. The projector input 226 is sent to the mobile device 202 and may be projected by the projector 210 into the scene. For example, an image of a cat 208 may be projected into the scene in a realistic manner taking into account the 3D model of the scene and the camera pose. The 3D model of the scene could be a computer aided design (CAD) model, or could be a model of the surfaces in the scene built up from images captured of the scene using a 3D model generation system 222. An example of a 3D model generation system which may be used is described in US patent application "Three-Dimensional Environment Reconstruction" Newcombe, Richard et al. published on Aug. 2, 2012 US20120194516. Other types of 3D model and 3D model generation systems may also be used.

An example where the downstream system 122 is a navigation system is now described with reference to FIG. 3. FIG. 3 has a plan view of a floor of an office 300 with various objects 310. A person 302 holding a mobile device 304 is walking along a corridor 306 in the direction of arrows 308. The mobile device 304 has one or more cameras 314, a camera pose tracker 316 and a map display 318. The mobile device 304 may be a smart phone or other mobile communications device as described with reference to FIG. 2 and which is able to communicate with a navigation system 322 in the cloud 320. The navigation system 322 receives the camera pose from the mobile device (where the user has consented to the disclosure of this information) and uses that information together with maps 324 of the floor of the office to calculate map display data to aid the person 302 in navigating the office floor. The map display data is sent to the mobile device and may be displayed at map display 318.

An example where the downstream system 122 is a robotic system is now described with reference to FIG. 3. A robot vehicle 312 moves along the corridor 306 and captures images using one or more cameras 326 on the robot vehicle. A camera pose tracker 328 at the robot vehicle is able to calculate pose of the camera(s) where the scene is already known to the robot vehicle.

FIG. 4 is a schematic diagram of an example decision forest comprising three decision trees: a first tree 400 (denoted tree  $\Psi_1$ ); a second tree 402 (denoted tree  $\Psi_2$ ); and a third tree 404 (denoted tree  $\Psi_3$ ). Each decision tree comprises a root node (e.g. root node 406 of the first decision tree 700), a plurality of internal nodes, called split nodes (e.g. split node 408 of the first decision tree 400), and a plurality of leaf nodes (e.g. leaf node 410 of the first decision tree 400).

## 6

In operation, each root and split node of each tree performs a binary test (or possibly an n-ary test) on the input data and based on the result directs the data to the left or right child node. The leaf nodes do not perform any action; they store accumulated scene coordinates (and optionally other information). For example, probability distributions may be stored representing the accumulated scene coordinates.

FIG. 5 is a flow diagram of a method of training a random decision forest to predict correspondences between image elements and scene coordinates. This is illustrated in the upper part of FIG. 5 above the dotted line in the region labeled "training". The lower part of FIG. 5 below the dotted line shows method steps at test time when the trained random decision forest is used to predict (or estimate) correspondences between image elements from an image of a scene and points in the scene's 3D world coordinate frame (scene coordinates).

A random decision forest is trained 502 to enable image elements to generate predictions of correspondences between themselves and scene coordinates. During training, labeled training images 500 of at least one scene, such as scene A, are used. For example, a labeled training image comprises, for each image element, a point in a scene's 3D world coordinate frame which the image element depicts. To obtain the labeled training images various different methods may be used to capture images 516 of scene A and record or calculate the pose of the camera for each captured image. Using this data a scene coordinate may be calculated indicating the world point depicted by an image element. To capture the images and record or calculate the associated camera pose, one approach is to carry out camera tracking from depth camera input 512. For example as described in US patent application "Real-time camera tracking using depth maps" Newcombe, Richard et al. published on Aug. 2, 2012 US20120196679. Another approach is to carry out dense reconstruction and camera tracking from RGB camera input 514. It is also possible to use a CAD model to generate synthetic training data. The training images themselves (i.e. not the label images) may be real or synthetic.

An example of the training process of box 502 is described below with reference to FIG. 6. The result of training is a trained random decision forest 504 for scene A (in the case where the training images were of scene A).

At test time an input image 508 of scene A is received and a plurality of image elements are selected from the input image. The image elements may be selected at random or in another manner (for example, by selecting such that spurious or noisy image elements are omitted). Each selected image element may be applied 506 to the trained decision forest to obtain predicted correspondences 510 between those image elements and points in the scene's 3D world coordinate frame.

FIG. 6 is a flow diagram of a method of training a random decision forest using images of a scene where image elements have labels indicating their corresponding scene coordinates. A training set of images of a scene is received 600 where the image elements have labels indicating the scene coordinate of the scene point they depict. A number of trees to be used in the decision forest is selected 602, for example, between 3 and 20 trees.

A decision tree from the decision forest is selected 604 (e.g. the first decision tree 600) and the root node 606 is selected 606. At least a subset of the image elements from each of the training images are then selected 608. For example, the image may be filtered to remove noisy or spurious image elements.



A random set of test parameters (also called weak learners) are then generated **610** for use by the binary test performed at the root node as candidate features. In one example, the binary test is of the form:  $\xi > f(x; \theta) > \tau$ , such that  $f(x; \theta)$  is a function applied to image element  $x$  with parameters  $\theta$ , and with the output of the function compared to threshold values  $\xi$  and  $\tau$ . If the result of  $f(x; \theta)$  is in the range between  $\xi$  and  $\tau$  then the result of the binary test is true. Otherwise, the result of the binary test is false. In other examples, only one of the threshold values  $\xi$  and  $\tau$  can be used, such that the result of the binary test is true if the result of  $f(x; \theta)$  is greater than (or alternatively less than) a threshold value. In the example described here, the parameter  $\theta$  defines a feature of the image.

A candidate function  $f(x; \theta)$  makes use of image information which is available at test time. The parameter  $\theta$  for the function  $f(x; \theta)$  is randomly generated during training. The process for generating the parameter  $\theta$  can comprise generating random spatial offset values in the form of a two or three dimensional displacement. The result of the function  $f(x; \theta)$  is then computed by observing the depth (or intensity value in the case of an RGB image and depth image pair) value for one or more test image elements which are displaced from the image element of interest  $x$  in the image by spatial offsets. The spatial offsets are optionally made depth invariant by scaling by  $1/\text{depth}$  of the image element of interest. Where RGB images are used without depth images the result of the function  $f(x; \theta)$  may be computed by observing the intensity value in a specified one of the red, green or blue color channel for one or more test image elements which are displaced from the image element of interest  $x$  in the image by spatial offsets.

The result of the binary test performed at a root node or split node determines which child node an image element is passed to. For example, if the result of the binary test is true, the image element is passed to a first child node, whereas if the result is false, the image element is passed to a second child node.

The random set of test parameters generated comprise a plurality of random values for the function parameter  $\theta$  and the threshold values  $\xi$  and  $\tau$ . In order to inject randomness into the decision trees, the function parameters  $\theta$  of each split node are optimized only over a randomly sampled subset  $\Theta$  of all possible parameters. This is an effective and simple way of injecting randomness into the trees, and increases generalization.

Then, every combination of test parameter may be applied **612** to each image element in the set of training images. In other words, available values for  $\theta$  (i.e.  $\theta_i \in \Theta$ ) are tried one after the other, in combination with available values of  $\xi$  and  $\tau$  for each image element in each training image. For each combination, criteria (also referred to as objectives) are calculated **614**. The combination of parameters that optimize the criteria is selected **614** and stored at the current node for future use.

In an example the objective is a reduction-in-variance objective expressed as follows:

$$Q(S_n, \theta) = V(S_n) - \sum_{d \in \{L, R\}} \frac{|S_n^d(\theta)|}{|S_n|} V(S_n^d(\theta))$$

Which may be expressed in words as the reduction in variance of the training examples at split node  $n$ , with weak learner parameters  $\theta$  equal to the variance of all the training

examples which reach that split node minus the sum of the variances of the training examples which reach the left and right child nodes of the split node. The variance may be calculated as:

$$V(S) = \frac{1}{|S|} \sum_{(p, m) \in S} \|m - \bar{m}\|_2^2$$

Which may be expressed in words as, the variance of a set of training examples  $S$  equals the average of the differences between the scene coordinates  $m$  and the mean of the scene coordinates in  $S$ .

As an alternative to a reduction-in-variance objective, other criteria can be used, such as logarithm of the determinant, or the continuous information gain.

It is then determined **616** whether the value for the calculated criteria is less than (or greater than) a threshold. If the value for the calculated criteria is less than the threshold, then this indicates that further expansion of the tree does not provide significant benefit. This gives rise to asymmetrical trees which naturally stop growing when no further nodes are beneficial. In such cases, the current node is set **618** as a leaf node. Similarly, the current depth of the tree is determined (i.e. how many levels of nodes are between the root node and the current node). If this is greater than a predefined maximum value, then the current node is set **618** as a leaf node. Each leaf node has scene coordinate predictions which accumulate at that leaf node during the training process as described below.

It is also possible to use another stopping criterion in combination with those already mentioned. For example, to assess the number of example image elements that reach the leaf. If there are too few examples (compared with a threshold for example) then the process may be arranged to stop to avoid overfitting. However, it is not essential to use this stopping criterion.

If the value for the calculated criteria is greater than or equal to the threshold, and the tree depth is less than the maximum value, then the current node is set **620** as a split node. As the current node is a split node, it has child nodes, and the process then moves to training these child nodes. Each child node is trained using a subset of the training image elements at the current node. The subset of image elements sent to a child node is determined using the parameters that optimized the criteria. These parameters are used in the binary test, and the binary test performed **622** on all image elements at the current node. The image elements that pass the binary test form a first subset sent to a first child node, and the image elements that fail the binary test form a second subset sent to a second child node.

For each of the child nodes, the process as outlined in blocks **610** to **622** of FIG. 6 are recursively executed **624** for the subset of image elements directed to the respective child node. In other words, for each child node, new random test parameters are generated **610**, applied **612** to the respective subset of image elements, parameters optimizing the criteria selected **614**, and the type of node (split or leaf) determined **616**. If it is a leaf node, then the current branch of recursion ceases. If it is a split node, binary tests are performed **622** to determine further subsets of image elements and another branch of recursion starts. Therefore, this process recursively moves through the tree, training each node until leaf nodes are reached at each branch. As leaf nodes are reached, the process waits **626** until the nodes in all branches have



been trained. Note that, in other examples, the same functionality can be attained using alternative techniques to recursion.

Once all the nodes in the tree have been trained to determine the parameters for the binary test optimizing the criteria at each split node, and leaf nodes have been selected to terminate each branch, then scene coordinates may be accumulated **628** at the leaf nodes of the tree. This is the training stage and so particular image elements which reach a given leaf node have specified scene coordinates known from the ground truth training data. A representation of the scene coordinates may be stored **630** using various different methods. For example by aggregating the scene coordinates or storing statistics representing the distribution of scene coordinates.

In some embodiments a multi-modal distribution is fitted to the accumulated scene coordinates. Examples of fitting a multi-model distribution include using expectation maximization (such as fitting a Gaussian mixture model); using mean shift mode detection; using any suitable clustering process such as k-means clustering, agglomerative clustering or other clustering processes. Characteristics of the clusters or multi-modal distributions are then stored rather than storing the individual scene coordinates. In some examples a handful of the samples of the individual scene coordinates may be stored.

A weight may also be stored for each cluster or mode. For example, a mean shift mode detection algorithm is used and the number of scene coordinates that reached a particular mode may be used as a weight for that mode. Mean shift mode detection is an algorithm that efficiently detects the modes (peaks) in a distribution defined by a Parzen window density estimator. In another example, the density as defined by a Parzen window density estimator may be used as a weight. A Parzen window density estimator (also known as a kernel density estimator) is a non-parametric process for estimating a probability density function, in this case of the accumulated scene coordinates. A Parzen window density estimator takes a bandwidth parameter which can be thought of as controlling a degree of smoothing.

In an example a sub-sample of the training image elements that reach a leaf are taken and input to a mean shift mode detection process. This clusters the scene coordinates into a small set of modes. One or more of these modes may be stored for example, according to the number of examples assigned to each mode.

Once the accumulated scene coordinates have been stored it is determined **632** whether more trees are present in the decision forest. If so, then the next tree in the decision forest is selected, and the process repeats. If all the trees in the forest have been trained, and no others remain, then the training process is complete and the process terminates **634**.

Therefore, as a result of the training process, one or more decision trees are trained using empirical training images. Each tree comprises a plurality of split nodes storing optimized test parameters, and leaf nodes storing associated scene coordinates or representations of aggregated scene coordinates. Due to the random generation of parameters from a limited subset used at each node, and the possible subsampled set of training data used in each tree, the trees of the forest are distinct (i.e. different) from each other.

The training process may be performed in advance of using the trained prediction system to identify scene coordinates for image elements of depth or RGB images of one or more known scenes. The decision forest and the opti-

mized test parameters may be stored on a storage device for use in identifying scene coordinates of image elements at a later time.

FIG. 7 illustrates a flowchart of a process for predicting scene coordinates in a previously unseen image (a depth image, an RGB image, or a pair of rectified depth and RGB images) using a decision forest that has been trained as described with reference to FIG. 6. Firstly, an unseen image is received **700**. An image is referred to as 'unseen' to distinguish it from a training image which has the scene coordinates already specified.

An image element from the unseen image is selected **702**. A trained decision tree from the decision forest is also selected **704**. The selected image element is pushed **706** through the selected decision tree, such that it is tested against the trained parameters at a node, and then passed to the appropriate child in dependence on the outcome of the test, and the process repeated until the image element reaches a leaf node. Once the image element reaches a leaf node, the accumulated scene coordinates (from the training stage) associated with this leaf node are stored **708** for this image element. In an example where the leaf node stores one or more modes of a distribution of scene coordinates, one or more of those modes are stored for this image element.

If it is determined **710** that there are more decision trees in the forest, then a new decision tree is selected **704**, the image element pushed **706** through the tree and the accumulated scene coordinates stored **708**. This is repeated until it has been performed for all the decision trees in the forest. The final prediction of the forest for an image element may be an aggregate of the scene coordinates obtained from the leaf found at each tree. Where one or more modes of a distribution of scene coordinates are stored at the leaves, the final prediction of the forest may be a union of the modes from the leaf found at each tree. Note that the process for pushing an image element through the plurality of trees in the decision forest can also be performed in parallel, instead of in sequence as shown in FIG. 7.

It is then determined **712** whether further unanalyzed image elements are to be assessed, and if so another image element is selected and the process repeated. The camera pose inference engine may be arranged to determine whether further unanalyzed image elements are to be assessed as described below with reference to FIG. 8.

FIG. 8 is a flow diagram of a method at a camera pose inference engine of using scene-coordinate-image element pairs to infer camera pose. As mentioned above the camera pose inference engine may use an energy optimization approach to find a camera pose which is a good fit to a plurality of image element—scene coordinate pairs predicted by the scene coordinate decision forest. In the case that depth images, or both depth and RGB images are used, an example energy function may be:

$$E(H) = \sum_{i \in I} \rho \left( \min_{m \in M_i} \|m - Hx_i\|_2 \right) = \sum_{i \in I} e_i(H)$$

Where  $i \in I$  is an image element index;  $\rho$  is a robust error function;  $m \in M_i$  represents the set of modes (3D locations in the scene's world space) predicted by the trees in the forest at image element  $p_i$ ; and  $x_i$  are the 3D coordinates in camera space corresponding to pixel  $p_i$  which may be obtained by back-projecting the depth image elements. The energy function may be considered as counting the number



## 11

of outliers for a given camera hypothesis  $H$ . The above notation uses homogeneous 3D coordinates.

In the case that RGB images are used without depth images the energy function may be modified by

$$E(H) = \sum_{i \in I} \rho(\min_{m \in M_i} \|\pi(KH^{-1}m - p_i)\|_2) = \sum_{i \in I} e_i(H)$$

where  $\rho$  is a robust error function,  $\pi$  projects from 3D to 2D image coordinates,  $K$  is a matrix that encodes the camera intrinsic parameters, and  $p_i$  is the 2D image element coordinate.

Note that  $E$ ,  $\rho$  and  $e_i$  may be separated out with different superscripts such as  $rgb/depth$  in the above equations.

In order to optimize the energy function an iterative process may be used to search for good camera pose candidates amongst a set of possible camera pose candidates. Samples of image element—scene coordinate pairs are taken and used to assess the camera pose candidates. The camera pose candidates may be refined or updated using a subset of the image element-scene coordinate pairs. By using samples of image element-scene coordinate pairs rather than each image element-scene coordinate pair from an image computation time is reduced without loss of accuracy.

An example iterative process which may be used at the camera pose inference engine is now described with reference to FIG. 8. A set of initial camera pose candidates or hypotheses is generated **800** by, for each camera pose candidate, selecting **802** three image elements from the input image (which may be a depth image, an RGB image or a pair of rectified depth and RGB images). The selection may be random or may take into account noise or missing values in the input image. It is also possible to pick pairs where the scene coordinate is more certain where certainty information is available from the forest. In some examples a minimum distance separation between the image elements may be enforced in order to improve accuracy. Each image element is pushed through the trained scene coordinate decision forest to obtain three scene coordinates. The three image element-scene coordinate pairs are used to compute **804** a camera pose using any suitable method such as the Kabsch algorithm also known as orthogonal Procrustes alignment which uses a singular value decomposition to compute the camera pose hypothesis. In some examples the set of initial camera pose candidates may include **820** one or more camera poses of previous frames where a stream of images is available. It may also include a camera pose predicted from knowledge of the camera's path.

For each camera pose hypothesis some inliers or outliers are computed **806**. Inliers and outliers are image element-scene coordinate pairs which are classified as either being consistent with a camera pose hypothesis or not. To compute inliers and outliers a batch  $B$  of image elements is sampled **808** from the input image and applied to the trained forest to obtain scene coordinates. The sampling may be random or may take into account noise or missing values in the input image. Each scene coordinate-image element pair may be classified **810** as an inlier or an outlier according to each of the camera pose hypotheses. For example, by comparing what the forest says the scene coordinate is for the image element and what the camera pose hypothesis says the scene coordinate is for the image element.

Optionally, one or more of the camera pose hypotheses may be discarded **812** on the basis of the relative number of

## 12

inliers (or outliers) associated with each hypothesis, or on the basis of a rank ordering by outlier count with the other hypotheses. In various examples the ranking or selecting hypotheses may be achieved by counting how many outliers each camera pose hypothesis has. Camera pose hypotheses with fewer outliers have a higher energy according to the energy function above.

Optionally, the remaining camera pose hypotheses may be refined **814** by using the inliers associated with each camera pose to recompute that camera pose (using the Kabsch algorithm mentioned above). For efficiency the process may store and update the means and covariance matrices used by the singular value decomposition.

The process may repeat **816** by sampling another batch  $B$  of image elements and so on until one or a specified number of camera poses remains or according to other criteria (such as the number of iterations).

The camera pose inference engine is able to produce an accurate camera pose estimate at interactive rates. This is achieved without an explicit 3D model of the scene having to be computed. A 3D model of the scene can be thought of as implicitly encoded in the trained random decision forest. Because the forest has been trained to work at any valid image element it is possible to sample image elements at test time. The sampling avoids the need to compute interest points and the expense of densely evaluating the forest.

FIG. 9 is a schematic diagram of the camera pose tracker of FIG. 1 where a 3D model **902** of the scene is available. For example the 3D model may be a CAD model or may be a dense reconstruction of the scene built up from depth images of the scene as described in US patent application "Three-dimensional environment reconstruction" Newcombe, Richard et al. published on Aug. 2, 2012 US20120194516. A pose refinement process **900** may be carried out to improve the accuracy of the camera pose **120**. The pose refinement process **900** may be an iterative closest point pose refinement as described in US patent application "Real-time camera tracking using depth maps" Newcombe, Richard et al. published on Aug. 2, 2012 US20120196679. In another example the pose refinement process **900** may seek to align depth observations from the mobile camera with surfaces of the 3D model of the scene in order to find an updated position and orientation of the camera which facilitates the alignment. This is described in U.S. patent application Ser. No. 13/749,497 filed on 24 Jan. 2013 entitled "Camera pose estimation for 3D reconstruction" Sharp et al.

The example shown in FIG. 9 has a camera pose tracker with one trained random decision forest rather than a plurality of trained random decision forests as in FIG. 1. This is intended to illustrate that a single forest may encapsulate a plurality of scenes by training the single forest using training data from those scenes. The training data comprises scene coordinates for image elements and also labels for image elements which identify a particular scene. Each sub-scene may be given a 3D sub-region of the full 3D world coordinate space and the forest may then be trained as described above. The camera pose tracker output may comprise an estimated camera pose and a scene so that the camera pose tracker is also able to carry out scene recognition. This enables the camera pose tracker to send data to a downstream system identifying which of a plurality of possible scenes the camera is in.

FIG. 10 illustrates various components of an exemplary computing-based device **1004** which may be implemented



as any form of a computing and/or electronic device, and in which embodiments of a camera pose tracker or object pose tracker may be implemented.

The computing-based device **1004** comprises one or more input interfaces **1002** arranged to receive and process input from one or more devices, such as user input devices (e.g. capture device **1008**, a game controller **1005**, a keyboard **1006**, a mouse **1007**). This user input may be used to control software applications, camera pose tracking or object pose tracking. For example, capture device **1008** may be a mobile depth camera arranged to capture depth maps of a scene. It may also be a fixed depth camera arranged to capture depth maps of an object. In another example, capture device **1008** comprises both a depth camera and an RGB camera. The computing-based device **1004** may be arranged to provide camera or object pose tracking at interactive rates.

The computing-based device **1004** also comprises an output interface **1010** arranged to output display information to a display device **1009** which can be separate from or integral to the computing device **1004**. The display information may provide a graphical user interface. In an example, the display device **1009** may also act as the user input device if it is a touch sensitive display device. The output interface **1010** may also output data to devices other than the display device, e.g. a locally connected printing device.

In some examples the user input devices **1005**, **1007**, **1008**, **1009** may detect voice input, user gestures or other user actions and may provide a natural user interface (NUI). This user input may be used to control a game or other application. The output interface **1010** may also output data to devices other than the display device, e.g. a locally connected printing device.

The input interface **1002**, output interface **1010**, display device **1009** and optionally the user input devices **1005**, **1007**, **1008**, **1009** may comprise NUI technology which enables a user to interact with the computing-based device in a natural manner, free from artificial constraints imposed by input devices such as mice, keyboards, remote controls and the like. Examples of NUI technology that may be provided include but are not limited to those relying on voice and/or speech recognition, touch and/or stylus recognition (touch sensitive displays), gesture recognition both on screen and adjacent to the screen, air gestures, head and eye tracking, voice and speech, vision, touch, gestures, and machine intelligence. Other examples of NUI technology that may be used include intention and goal understanding systems, motion gesture detection systems using depth cameras (such as stereoscopic camera systems, infrared camera systems, rgb camera systems and combinations of these), motion gesture detection using accelerometers/gyroscopes, facial recognition, 3D displays, head, eye and gaze tracking, immersive augmented reality and virtual reality systems and technologies for sensing brain activity using electric field sensing electrodes (EEG and related methods).

Computer executable instructions may be provided using any computer-readable media that is accessible by computing based device **1004**. Computer-readable media may include, for example, computer storage media such as memory **1012** and communications media. Computer storage media, such as memory **1012**, includes volatile and non-volatile, removable and non-removable media implemented in any method or technology for storage of information such as computer readable instructions, data structures, program modules or other data. Computer storage media includes, but is not limited to, RAM, ROM, EPROM, EEPROM, flash memory or other memory technology, CD-

ROM, digital versatile disks (DVD) or other optical storage, magnetic cassettes, magnetic tape, magnetic disk storage or other magnetic storage devices, or any other non-transmission medium that can be used to store information for access by a computing device.

In contrast, communication media may embody computer readable instructions, data structures, program modules, or other data in a modulated data signal, such as a carrier wave, or other transport mechanism. As defined herein, computer storage media does not include communication media. Therefore, a computer storage medium should not be interpreted to be a propagating signal per se. Propagated signals may be present in a computer storage media, but propagated signals per se are not examples of computer storage media. Although the computer storage media (memory **1012**) is shown within the computing-based device **1004** it will be appreciated that the storage may be distributed or located remotely and accessed via a network or other communication link (e.g. using communication interface **1013**).

Computing-based device **1004** also comprises one or more processors **1000** which may be microprocessors, controllers or any other suitable type of processors for processing computing executable instructions to control the operation of the device in order to provide real-time camera tracking. In some examples, for example where a system on a chip architecture is used, the processors **1000** may include one or more fixed function blocks (also referred to as accelerators) which implement a part of the method of real-time camera tracking in hardware (rather than software or firmware).

Platform software comprising an operating system **1014** or any other suitable platform software may be provided at the computing-based device to enable application software **1016** to be executed on the device. Other software than may be executed on the computing device **1004** comprises: camera/object pose tracker **1018** which comprises a pose inference engine. A trained support vector machine regression system may also be provided and/or a trained Gaussian process regression system. A data store **1020** is provided to store data such as previously received images, camera pose estimates, object pose estimates, trained random decision forests registration parameters, user configurable parameters, other parameters, 3D models of scenes, game state information, game metadata, map data and other data.

The term 'computer' or 'computing-based device' is used herein to refer to any device with processing capability such that it can execute instructions. Those skilled in the art will realize that such processing capabilities are incorporated into many different devices and therefore the terms 'computer' and 'computing-based device' each include PCs, servers, mobile telephones (including smart phones), tablet computers, set-top boxes, media players, games consoles, personal digital assistants and many other devices.

The methods described herein may be performed by software in machine readable form on a tangible storage medium e.g. in the form of a computer program comprising computer program code means adapted to perform all the steps of any of the methods described herein when the program is run on a computer and where the computer program may be embodied on a computer readable medium. Examples of tangible storage media include computer storage devices comprising computer-readable media such as disks, thumb drives, memory etc. and do not include propagated signals. Propagated signals may be present in a tangible storage media, but propagated signals per se are not examples of tangible storage media. The software can be suitable for execution on a parallel processor or a serial



processor such that the method steps may be carried out in any suitable order, or simultaneously.

This acknowledges that software can be a valuable, separately tradable commodity. It is intended to encompass software, which runs on or controls “dumb” or standard hardware, to carry out the desired functions. It is also intended to encompass software which “describes” or defines the configuration of hardware, such as HDL (hardware description language) software, as is used for designing silicon chips, or for configuring universal programmable chips, to carry out desired functions.

Those skilled in the art will realize that storage devices utilized to store program instructions can be distributed across a network. For example, a remote computer may store an example of the process described as software. A local or terminal computer may access the remote computer and download a part or all of the software to run the program. Alternatively, the local computer may download pieces of the software as needed, or execute some software instructions at the local terminal and some at the remote computer (or computer network). Those skilled in the art will also realize that by utilizing conventional techniques known to those skilled in the art that all, or a portion of the software instructions may be carried out by a dedicated circuit, such as a DSP, programmable logic array, or the like.

Any range or device value given herein may be extended or altered without losing the effect sought, as will be apparent to the skilled person.

Although the subject matter has been described in language specific to structural features and/or methodological acts, it is to be understood that the subject matter defined in the appended claims is not necessarily limited to the specific features or acts described above. Rather, the specific features and acts described above are disclosed as example forms of implementing the claims.

It will be understood that the benefits and advantages described above may relate to one embodiment or may relate to several embodiments. The embodiments are not limited to those that solve any or all of the stated problems or those that have any or all of the stated benefits and advantages. It will further be understood that reference to ‘an’ item refers to one or more of those items.

The steps of the methods described herein may be carried out in any suitable order, or simultaneously where appropriate. Additionally, individual blocks may be deleted from any of the methods without departing from the spirit and scope of the subject matter described herein. Aspects of any of the examples described above may be combined with aspects of any of the other examples described to form further examples without losing the effect sought.

The term ‘comprising’ is used herein to mean including the method blocks or elements identified, but that such blocks or elements do not comprise an exclusive list and a method or apparatus may contain additional blocks or elements.

It will be understood that the above description is given by way of example only and that various modifications may be made by those skilled in the art. The above specification, examples and data provide a complete description of the structure and use of exemplary embodiments. Although various embodiments have been described above with a certain degree of particularity, or with reference to one or more individual embodiments, those skilled in the art could make numerous alterations to the disclosed embodiments without departing from the spirit or scope of this specification.

The invention claimed is:

1. A device-implemented method of calculating pose of an entity comprising:
  - receiving an unseen image where the unseen image is either: of the entity, or of a scene;
  - applying a plurality of image elements of the unseen image to each of a plurality of trained random forests; based at least on applying the plurality of image elements, receiving a plurality of image element-coordinate pairs, each of the plurality of image element-coordinate pairs comprising an input image element and a corresponding output coordinate comprising either entity coordinates or scene coordinates;
  - determining whether a pose of the entity has been calculated;
  - based on a determination that the pose of the entity has not been calculated, using the unseen image to calculate an initial pose of the entity without use of a three-dimensional model by applying the plurality of image element-coordinate pairs to a pose inference engine;
  - confirming the initial pose is accurate by:
    - sampling a set of image elements from the unseen image;
    - applying the set of image elements to the plurality of trained random forests to obtain a set of scene coordinates;
    - comparing what the plurality of trained random forests indicate the set of scene coordinates are for the set of image elements to what the initial pose indicates the set of scene coordinates are for the image elements; and
    - classifying each scene coordinate in the set of scene coordinates as an inlier or an outlier according to the initial pose based on the comparing; and
    - based on a number of inliers and based on a rank ordering by outlier count of the initial pose and other hypotheses of the pose of the entity, confirming that the initial pose is accurate; and
  - generating map display data based at least in part on the calculated initial pose of the entity.
2. The device-implemented method as claimed in claim 1, wherein the entity is a mobile camera and the initial pose of the mobile camera is calculated without using a depth image.
3. The device-implemented method as claimed in claim 1, wherein the entity is an object and the initial pose of the object is calculated using only the unseen image captured by a fixed camera.
4. The device-implemented method as claimed in claim further comprising calculating the pose of the entity as parameters having six degrees of freedom, three indicating rotation of the entity and three indicating position of the entity.
5. The device-implemented method as claimed in claim 1, wherein the unseen image is an image in which scene coordinates are unknown.
6. The device-implemented method as claimed in claim 1, wherein no other image besides the unseen image is used to calculate the initial pose of the entity.
7. The device-implemented method as claimed in claim further comprising a machine learning system comprising the plurality of trained random forests and the method comprises applying the image elements of the unseen image to the plurality of trained random forests, each random forest having been trained using images from a different one of a plurality of scenes, obtaining scene coordinates from the applied image elements, and calculating the initial pose of the entity from the scene coordinates.



17

8. The device-implemented method as claimed in claim further comprising calculating the initial pose by searching amongst a set of possible pose candidates and using samples of associations between image elements and points to assess the set of possible pose candidates.

9. The device-implemented method as claimed in claim further comprising receiving at the processor, a stream of images, and calculating the initial pose by searching amongst a set of possible pose candidates which includes a pose calculated from another image in the stream of images.

10. The device-implemented method as claimed in claim 1, at least partially carried out using hardware logic selected from any one or more of: a field-programmable gate array, a program-specific integrated circuit, a program-specific standard product, a system-on-a-chip, a complex programmable logic device, a graphics processing unit.

11. The device-implemented method as claimed in claim wherein the entity is a mobile camera and the pose of the mobile camera is calculated, the method comprising accessing a 3D model of the scene and refining the pose of the mobile camera using the accessed 3D model.

12. A pose tracker comprising:

a processor arranged to receive an unseen image of a scene captured by a mobile camera;

the processor arranged to:

apply a plurality of image elements of the unseen image to each of a plurality of trained random forests;

based at least on applying the plurality of image elements, receiving a plurality of image element-coordinate pairs, each of the plurality of image element-coordinate pairs comprising an input image element and a corresponding output coordinate comprising either entity coordinates or scene coordinates; and

determine whether a pose of the mobile camera has been calculated;

using the unseen image to calculate an initial pose of the mobile camera without use of a three-dimensional model by applying the plurality of image element-coordinate pairs to a pose inference engine;

confirming the initial pose is accurate by:

sampling a set of image elements from the unseen image;

applying the set of image elements to the plurality of trained forests to obtain a set of scene coordinates;

comparing what the plurality of trained forests indicate the set of scene coordinates are for the set of image elements to what the initial pose indicates the set of scene coordinates are for the image elements; and

classifying each scene coordinate in the set of scene coordinates as an inlier or an outlier according to the initial pose based on the comparing; and

based on a number of inliers and based on a rank ordering by entity, confirming that the initial pose is accurate; and

generate map display data based at least in part on the calculated initial pose of the mobile camera.

13. The pose tracker as claimed in claim 12, wherein the processor is arranged to apply only a subsample of the image elements of the unseen image to a trained machine learning system.

14. The pose tracker as claimed in claim 12, wherein the pose inference engine is arranged to calculate the initial pose by searching amongst a set of possible pose candidates and

18

using samples of associations between image elements and points in scene coordinates to assess the set of possible pose candidates.

15. The pose tracker as claimed in claim 12, wherein the processor is arranged to receive a stream of images, and wherein the pose inference engine is arranged to calculate the initial pose by searching amongst a set of possible pose candidates which includes a pose calculated from another image in the stream of images.

16. The pose tracker as claimed in claim 12 at least partially implemented using hardware logic selected from any one or more of: a field-programmable gate array, a program-specific integrated circuit, a program-specific standard product, a system-on-a-chip, a complex programmable logic device, a graphics processing unit.

17. One or more computer-readable storage devices comprising computer-executable instructions that, when executed by one or more processors, cause the one or more processors to perform operations comprising:

receiving an unseen image where the unseen image is either: of an entity or of a scene;

applying a plurality of image elements of the unseen image to each of a plurality of trained random forests;

based at least on applying the plurality of image elements, receiving a plurality of image element-coordinate pairs,

each of the plurality of image element-coordinate pairs comprising an input image element and a corresponding output coordinate comprising either entity coordinates or scene coordinates;

determining whether a pose of the entity has been calculated;

based on a determination that the pose of the entity has not been calculated, using the unseen image to calculate an initial pose of the entity without use of a three-dimensional model by applying the plurality of image element-coordinate pairs to a pose inference engine;

confirming the initial pose is accurate by:

sampling a set of image elements from the unseen image;

applying the set of image elements to the plurality of trained random forests to obtain a set of scene coordinates;

comparing what the plurality of trained random forests indicate the set of scene coordinates are for the set of image elements to what the initial pose indicates the set of scene coordinates are for the image elements; and

classifying each scene coordinate in the set of scene coordinates as an inlier or an outlier according to the initial pose based on the comparing; and

based on a number of inliers and based on a rank ordering by outlier count of the initial pose and other hypotheses of the pose of the entity, confirming that the initial pose is accurate; and

generating map display data based at least in part on the calculated initial pose of the entity.

18. The one or more computer-readable storage devices according to claim 17, where the entity is a mobile camera and the initial pose of the mobile camera is calculated without using a depth image.

19. The one or more computer-readable storage devices according to claim 17, where the entity is an object and the initial pose of the object is calculated using only the unseen image captured by a fixed camera.



20. The one or more computer-readable storage devices according to claim 17, wherein the unseen image is an image in which scene coordinates are unknown.

\* \* \* \* \*